

P

SEARCH REQUEST FORM
Scientific and Technical Information Center

Access DB#

51641

Requester's Full Name: _____ Examiner #: _____ Date: _____
Art Unit: _____ Phone Number 30 _____ Serial Number: _____
Mail Box and Bldg/Room Location: _____ Results Format Preferred (circle): PAPER DISK E-MAIL

If more than one search is submitted, please prioritize searches in order of need.

Please provide a detailed statement of the search topic, and describe as specifically as possible the subject matter to be searched. Include the elected species or structures, keywords, synonyms, acronyms, and registry numbers, and combine with the concept or utility of the invention. Define any terms that may have a special meaning. Give examples or relevant citations, authors, etc, if known. Please attach a copy of the cover sheet, pertinent claims, and abstract.

Title of Invention: _____

Inventors (please provide full names): _____

Earliest Priority Filing Date: _____

**For Sequence Searches Only* Please include all pertinent information (parent, child, divisional, or issued patent numbers) along with the appropriate serial number.*

STAFF USE ONLY

	Type of Search	Vendors and cost where applicable
Searcher: <u>Toby Port</u>	NA Sequence (#) <u>1</u>	STN _____
Searcher Phone #: <u>308-3534</u>	AA Sequence (#) <u>2</u>	Dialog _____
Searcher Location: _____	Structure (#) _____	Questel/Orbit _____
Date Searcher Picked Up: <u>9/26</u>	Bibliographic _____	Dr.Link _____
Date Completed: <u>9/26</u>	Litigation _____	Lexis/Nexis _____
Searcher Prep & Review Time: <u>10</u>	Fulltext _____	Sequence Systems <u>CS</u>
Clerical Prep Time: _____	Patent Family _____	WWW/Internet _____
Online Time: <u>10</u>	Other _____	Other (specify) _____

PTO-1590 (1-2000)

BEST AVAILABLE COPY

THIS PAGE BLANK (USPTO)

```
US-09-615-285-1  
Title: 1738  
Perfect score: 1 ggcgagagcggaagcggag.....ctgcgaacaa  
Sequence: IDENTITY_NUC Gapext 1.0  
Gapop 10.0 , Gapext residues  
Scoring table: Gapop 1.0 segs, 7733874598 residues  
Parameters:
```

```
Searched:
Total number of hits satisfying criteria: 1
Minimum DB seq length: 2000000000
Maximum DB seq length: 2000000000
Minimum Match 0%
Maximum Match 100%
Post-Processing: Minimum Match 100%
                  Maximum Match 100%
                  Listing first 45 summaries
```

Database :

```

1: GenEmb: *
2:   qb_ba1: *
3:   qb_ba2: *
4:   qb_ba3: *
5:   qb_in1: *
6:   qb_in2: *
7:   qb_in3: *
8:   qb_cm1: *
9:   qb_cm2: *
10:  qb_cm3: *
11:  qb_ph1: *
12:  qb_ph2: *
13:  qb_ph3: *
14:  qb_p1: *
15:  qb_p2: *
16:  qb_p3: *
17:  em_ba1: *
18:  em_ba2: *
19:  em_ba3: *
20:  em_hg1: *
21:  em_hg2: *
22:  em_hg3: *
23:  em_hg4: *
24:  em_hg5: *
25:  em_hg6: *
26:  em_hg7: *
27:  em_hg8: *
28:  em_hg9: *
29:  em_hg10: *
30:  em_hg11: *
31:  em_hg12: *
32:  em_hg13: *
33:  em_hg14: *
34:  em_hg15: *
35:  em_hg16: *
36:  em_hg17: *
37:  em_hg18: *
38:  em_hg19: *
39:  em_hg20: *
40:  em_in1: *
41:  em_in2: *
42:  em_in3: *
43:  em_or: *

```

44: em.ov.*
45: em.pat.*
46: em.ph.*
47: em.pl.*
48: em.ro.*
49: em.sts.*
50: em.sy.*
51: em.un.*
52: qb.sts1.*
53: qb.sts2.*
54: qb.sts3.*
55: qb.sts.*
56: qb.sy.*
57: qb.un.*
58: qb.v1.*
59: qb.v12.*
60: qb.hcg1.*
61: qb.hcg2.*
62: qb.hcg3.*
63: qb.hcg4.*
64: qb.hcg5.*
65: qb.hcg6.*
66: qb.hcg7.*
67: qb.hcg8.*
68: qb.hcg9.*
69: qb.hcg10.*
70: qb.hcg11.*
71: qb.hcg12.*
72: qb.hcg13.*
73: qb.hcg14.*
74: qb.hcg15.*
75: qb.hcg16.*
76: qb.hcg17.*
77: qb.hcg18.*
78: qb.hcg19.*
79: qb.hcg20.*
80: qb.hcg21.*
81: qb.hcg22.*
82: qb.hcg23.*
83: qb.hcg24.*
84: qb.hcg25.*
85: qb.pr1.*
86: qb.pr2.*
87: qb.pr3.*
88: qb.pr4.*
89: qb.pr5.*
90: qb.pr6.*
91: qb.pr7.*
92: qb.pr8.*
93: qb.pr9.*
94: qb.pr10.*
95: qb.pr11.*
96: qb.pr12.*
97: qb.pr13.*
98: qb.pr14.*
99: qb.pr15.*
100: qb.pr16.*
101: qb.pr17.*
102: qb.pr18.*
103: qb.pr19.*
104: qb.pr20.*
105: qb.pr21.*
106: qb.pr22.*
107: qb.pr23.*
108: qb.pr24.*
109: qb.pr25.*
110: qb.pr26.*
111: qb.pr27.*
112: qb.pr28.*
113: qb.pr29.*
114: qb.pr30.*
115: qb.pr31.*
116: qb.pr32.*
117: qb.pr33.*
118: qb.pr34.*
119: qb.pr35.*
120: qb.pr36.*
121: qb.pr37.*
122: qb.pr38.*
123: qb.pr39.*
124: qb.pr40.*
125: qb.pr41.*
126: qb.pr42.*
127: qb.pr43.*
128: qb.pr44.*
129: qb.pr45.*
130: qb.pr46.*
131: qb.pr47.*
132: qb.pr48.*
133: qb.pr49.*
134: qb.pr50.*
135: qb.pr51.*
136: qb.pr52.*
137: qb.pr53.*
138: qb.pr54.*
139: qb.pr55.*
140: qb.pr56.*
141: qb.pr57.*
142: qb.pr58.*
143: qb.pr59.*
144: qb.pr60.*
145: qb.pr61.*
146: qb.pr62.*
147: qb.pr63.*
148: qb.pr64.*
149: qb.pr65.*
150: qb.pr66.*
151: qb.pr67.*
152: qb.pr68.*
153: qb.pr69.*
154: qb.pr70.*
155: qb.pr71.*
156: qb.pr72.*
157: qb.pr73.*
158: qb.pr74.*
159: qb.pr75.*
160: qb.pr76.*
161: qb.pr77.*
162: qb.pr78.*
163: qb.pr79.*
164: qb.pr80.*
165: qb.pr81.*
166: qb.pr82.*
167: qb.pr83.*
168: qb.pr84.*
169: qb.pr85.*
170: qb.pr86.*
171: qb.pr87.*
172: qb.pr88.*
173: qb.pr89.*
174: qb.pr90.*
175: qb.pr91.*
176: qb.pr92.*
177: qb.pr93.*
178: qb.pr94.*
179: qb.pr95.*
180: qb.pr96.*
181: qb.pr97.*
182: qb.pr98.*
183: qb.pr99.*
184: qb.pr100.*
185: qb.pr101.*
186: qb.pr102.*
187: qb.pr103.*
188: qb.pr104.*
189: qb.pr105.*
190: qb.pr106.*
191: qb.pr107.*
192: qb.pr108.*
193: qb.pr109.*
194: qb.pr110.*
195: qb.pr111.*
196: qb.pr112.*
197: qb.pr113.*
198: qb.pr114.*
199: qb.pr115.*
200: qb.pr116.*
201: qb.pr117.*
202: qb.pr118.*
203: qb.pr119.*
204: qb.pr120.*
205: qb.pr121.*
206: qb.pr122.*
207: qb.pr123.*
208: qb.pr124.*
209: qb.pr125.*
210: qb.pr126.*
211: qb.pr127.*
212: qb.pr128.*
213: qb.pr129.*
214: qb.pr130.*
215: qb.pr131.*
216: qb.pr132.*
217: qb.pr133.*
218: qb.pr134.*
219: qb.pr135.*
220: qb.pr136.*
221: qb.pr137.*
222: qb.pr138.*
223: qb.pr139.*
224: qb.pr140.*
225: qb.pr141.*
226: qb.pr142.*
227: qb.pr143.*
228: qb.pr144.*
229: qb.pr145.*
230: qb.pr146.*
231: qb.pr147.*
232: qb.pr148.*
233: qb.pr149.*
234: qb.pr150.*
235: qb.pr151.*
236: qb.pr152.*
237: qb.pr153.*
238: qb.pr154.*
239: qb.pr155.*
240: qb.pr156.*
241: qb.pr157.*
242: qb.pr158.*
243: qb.pr159.*
244: qb.pr160.*
245: qb.pr161.*
246: qb.pr162.*
247: qb.pr163.*
248: qb.pr164.*
249: qb.pr165.*
250: qb.pr166.*
251: qb.pr167.*
252: qb.pr168.*
253: qb.pr169.*
254: qb.pr170.*
255: qb.pr171.*
256: qb.pr172.*
257: qb.pr173.*
258: qb.pr174.*
259: qb.pr175.*
260: qb.pr176.*
261: qb.pr177.*
262: qb.pr178.*
263: qb.pr179.*
264: qb.pr180.*
265: qb.pr181.*
266: qb.pr182.*
267: qb.pr183.*
268: qb.pr184.*
269: qb.pr185.*
270: qb.pr186.*
271: qb.pr187.*
272: qb.pr188.*
273: qb.pr189.*
274: qb.pr190.*
275: qb.pr191.*
276: qb.pr192.*
277: qb.pr193.*
278: qb.pr194.*
279: qb.pr195.*
280: qb.pr196.*
281: qb.pr197.*
282: qb.pr198.*
283: qb.pr199.*
284: qb.pr200.*
285: qb.pr201.*
286: qb.pr202.*
287: qb.pr203.*
288: qb.pr204.*
289: qb.pr205.*
290: qb.pr206.*
291: qb.pr207.*
292: qb.pr208.*
293: qb.pr209.*
294: qb.pr210.*
295: qb.pr211.*
296: qb.pr212.*
297: qb.pr213.*
298: qb.pr214.*
299: qb.pr215.*
300: qb.pr216.*
301: qb.pr217.*
302: qb.pr218.*
303: qb.pr219.*
304: qb.pr220.*
305: qb.pr221.*
306: qb.pr222.*
307: qb.pr223.*
308: qb.pr224.*
309: qb.pr225.*
310: qb.pr226.*
311: qb.pr227.*
312: qb.pr228.*
313: qb.pr229.*
314: qb.pr230.*
315: qb.pr231.*
316: qb.pr232.*
317: qb.pr233.*
318: qb.pr234.*
319: qb.pr235.*
320: qb.pr236.*
321: qb.pr237.*
322: qb.pr238.*
323: qb.pr239.*
324: qb.pr240.*
325: qb.pr241.*
326: qb.pr242.*
327: qb.pr243.*
328: qb.pr244.*
329: qb.pr245.*
330: qb.pr246.*
331: qb.pr247.*
332: qb.pr248.*
333: qb.pr249.*
334: qb.pr250.*
335: qb.pr251.*
336: qb.pr252.*
337: qb.pr253.*
338: qb.pr254.*
339: qb.pr255.*
340: qb.pr256.*
341: qb.pr257.*
342: qb.pr258.*
343: qb.pr259.*
344: qb.pr260.*
345: qb.pr261.*
346: qb.pr262.*
347: qb.pr263.*
348: qb.pr264.*
349: qb.pr265.*
350: qb.pr266.*
351: qb.pr267.*
352: qb.pr268.*
353: qb.pr269.*
354: qb.pr270.*
355: qb.pr271.*
356: qb.pr272.*
357: qb.pr273.*
358: qb.pr274.*
359: qb.pr275.*
360: qb.pr276.*
361: qb.pr277.*
362: qb.pr278.*
363: qb.pr279.*
364: qb.pr280.*
365: qb.pr281.*
366: qb.pr282.*
367: qb.pr283.*
368: qb.pr284.*
369: qb.pr285.*
370: qb.pr286.*
371: qb.pr287.*
372: qb.pr288.*
373: qb.pr289.*
374: qb.pr290.*
3

Result No.	Score	% Match Length DB	ID	Description
1	1738	100.0	AE720487	AE720487 Homo sapi
2	1655.2	95.2	HSU75329	U75329 Human scap
3	1654.2	95.2	AX011973	AX011973 Sequencia
4	1474.6	84.8	AF113353	AF113353 Homo sapi
5	974	56.6	AF119352	AF119352 Mus musc
6	972.6	56.6	AF113550	AF113550 Mus musc
7	969	55.8	AE243500	AE243500 Mus musc
8	295.6	17.0	AB038497	AB038497 Xenopus l

[illegible]

QY	181	aaccctctccgcgaagcccaactggtgtgtcccaactgtctaaagagtgatgtgtccgtccag	243
Db	183	AACCCCTATCCCGACAGACGCCACATGTGTGCCACATGTCTACAGAGGTGCATCCGGCTAG	242
QY	241	tactaacccgtcccccgtgtcccagtaagcccccgaaggtctctaaagagctttcaaaccc	300
Db	243	TACTAACCCGTCGCCCGTGTCCCGCACTAGCGCCCGAGGGTCTGTACGACGGCTTCCAAACCC	302
QY	301	gtcgtctcagcagcagcccaaaatccccaatccggagacagtgtgcacctcaagaataa	360
Db	303	GTCTCTGACAGCCAGCCCAATATCCCATCGGGAGCACTGTGCACCTCAAAAGATAAGAA	362
QY	361	gcactgtgatatcaacttgaacctgtgggaacctctgtgtggaaactgtgctgtgscgtgtgc	420
Db	363	GCACGTGTGATACCTTGACCTGTGGGGACCTTCCTCTGTGGAGCTGTGGCTGGCGCTTGGC	422
QY	421	ctaaccttggaaatttaaggagcagaagtgctcaaacctctggatlaagatgtgcacttcca	480
Db	423	CTACTCTGGAATTCATGTGGCAGCAAGTCTCAACCTGTGGATAGAGTGCAGCTCTCA	482
QY	481	gttaccctgtatcaaacccctctaaactgtgtgtgaatggcgtgtcaacatgtccccgcggggag	540
Db	483	GGTACCTGCATCAACCCCTCTTAACGTGGTGTATGGCGTGTACACTCTCCCGCGGGAGAG	542
QY	541	gaagagaatccgtgtgtgtgtgctctaaaggaccacaactcatccctcaagtgtaactaact	600
Db	543	GACGAGAAATCGGTGTGTGTGCCCTCTTAAGAACCAAACTTCATCTTCAAGTGTATCATCT	602
QY	601	cagaggaagatcctcgtgcacctgtgtgtccaagaagcctgtgaacgagaactacggtgcggcg	660
Db	603	CAGAGGAAGTCTCGGACACCTGTGTGCCAAGCGAGCTGAAGAGAACTACGGCGCGGCG	662
QY	661	gccttcagggagatgtggctataagaataattttactctaaaggaagatgtgtgaac	720


```
D 663 GCCTGCAGGACATGGGCTATAGAAATATTTTACTCTAGCCAGGAATAGTGATGAC 722
Q 721 agcggatccacccagcttattgaactgacacaaagtccggcaatctgatctataa 780
D 723 AGCGGATCCACGAGCTTTATGAAATGAAACACAAAGTCCGGCAATGTGATATCTATAA 782
Q 781 aaacttacacagtgatgcctcttctcaaaagcagtgattcttctaaccttataacc 840
D 783 AAACCTTACACAGTATGCTTCTTTCTTCAAAAGCAGTGGTTTCTTTACGCTTATAGCC 842
Q 841 tgcggggtcaacttgaactcaagccgcagcagagatttggcgcgagcagcgcgtc 900
D 843 TCGGGGGTCAACTTGAATCAACCCGCCAGAGAGATGTGGCGCGGAGAGCGCGCTC 902
Q 901 ccgggggcttgccctgcaagtcaacccctgcaagcttcaagcagcttgcggagcg 960
D 903 CCGGGGGCTGGCCCTGGAGGTCACCTGCACAGTCCAGAACTCCACGTGTCGGAGGC 962
Q 961 tccatcatcaccccccagtgatcgtgacacgcccgcactgcgtggaanaaccttaac 1020
D 963 TCCATCATCACCCCGGAGTGCATGCTGACAGCCGCCACTGCGTGGAAAACTCTTTAAC 1022
Q 1021 aatccatggcattgagcggcatttgcgggattttagaacaacttctatgtctatgga 1080
D 1023 AATCCATGGCATTGGACGCATTTGGGGGATTTGAGACAAATCTTTCAATGTTCTATGGA 1082
Q 1081 gccggataccagtgagaaaaagtatttctatccaatattagctccaaagccaagaac 1140
D 1083 GCGGGATACCAAGTAAAGAAAGATTTCTCAACCAATTAATGACCAAGACCAAGAAC 1142
Q 1141 aatgaatcgcgtgtagaagctgcagaaagcctcgtacttcaagcagctagtgaaacca 1200
D 1143 AATGACATTGGCTGTATGAAAGTGCAGAAAGCCCTCTACTTTTCAACAGCACTAGTAAACCA 1202
Q 1201 gtcgtctgcacaaaccaaagcaltgctgcagccagaaagcctcgtcgtgatttccggg 1260
D 1203 GTGTGCTCTCCCAACCCAGCATGATGATGTCAGACCCAGAACAGCTGTGAGATTTCCGGG 1262
Q 1261 tggggggcaccagagagaaaggaagcctcaagaagtctgaagctgcgcaagtgctt 1320
D 1263 TGGGGGGCCACCGAGAGAAAGGAGACCTCAGAAAGTCTGAAACCTCCCAAGGTGCTT 1322
Q 1321 ctcattgagacacagatgacacagacagatatgtctatgaacaacttgatcaaccagcc 1380
D 1323 CTCATTGACACACAGATGCAACACACATATGCTATGACACACTGATCACACACAGCC 1382
Q 1381 atgattctgtccggtctcctcgaagggaacgttcgattcttgcagggtgacagtggagg 1440
D 1383 ATGATCTGTGCCGGCTTCTCGACAGGGGAACGTGATTTTGGCCAGGGTGACATGGAGGG 1442
Q 1441 cctctggttacttcgaagaacaatattctggtgctgataaggatacaagctggggcttct 1500
D 1443 CCTGTGGTACTTTCGAAAGCAATATCTGTGGCTGATAGGGGATPCCAAGTGGGGTTCT 1502
Q 1501 ggcgtgccaagacttaagcagcagagtgctacgggaatgctatgattatcaaggacgtg 1560
D 1503 GCGGTGCGCAAAAGCTTACAGACAGAGAGTGTACGGGAATGTGATTTACAGGAGCTGG 1562
Q 1561 attatcgaacaaatgaaggcagacaggttaatacgaatggtcttcgtccttgcagctgctt 1620
D 1563 ATTATTCGAACAATGAGGCAACAGCGCTAATCCATGATGCTTCGTTGACGCTGTTT 1622
Q 1621 ttcagaagaaacaatgggtgtgttctgcttccccgtgcatgatttactcttagagatgat 1680
D 1623 TACAAAGAAAACAATGGGCTGTTGCTTCCCGGATGATTTACTCTTAAAGATGAT 1682
Q 1681 tcagaagctactcatatttataacagtgaaactgtctgcaaaaaaataaaaaaa 1738
D 1683 TCAGAGGTACTCTCAATTTTATTAAACATGAACTTGTGTGGCAAAAAAATAAAAAA 1740
```

RESULT 2
HSU75329

```
LOCUS HSU75329 2479 bp mRNA PRI 10-OCT-1997
DEFINITION Human serine protease mRNA, complete cds.
ACCESSION U75329
VERSION 075329.1 GI:2507612
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
AUTHORS Paoloni-Giacobino,A., Chen,H., Peltsch,M.C., Rossier,C. and
Antonarakis,S.E.
TITLE Cloning of the TMPRSS2 gene, which encodes a novel serine protease
with transmembrane, ID1RA, and SRCH domains and maps to 21q22.3
JOURNAL Genomics 44 (3), 309-320 (1997)
MEDLINE 97468144
REFERENCE 2 (bases 1 to 2479)
AUTHORS Paoloni-Giacobino,A., Chen,H. and Antonarakis,S.E.
TITLE Direct Submission
JOURNAL Submitted (17-OCT-1996) Medical Genetics, University of Geneva
Medical School, 1 Michel-Servet Street, Geneva 1211, Switzerland
FEATURES
source
1..2479
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="21"
/map="21q22.3"
57..1535
/codon_start=1
/product="serine protease"
/protein_id="AAC51784.1"
/db_xref="GI:2507613"
/translation="MALNSGSEPAIGPIYENHGYDENPYPAQPTVPVPEVPEVHPAY
YPSVPQYAPRVLTQASNPVCTOPKSPGVCTSKTKALCITLTLFLTGALAA
GLMKFMGSKCSNGSIEDSGCTINPSMDCGSHCHGDEENECVRJLYGNFTIOM
YSSORSMHPVCDODMNEYGRACRDMQYKKNFYSSOIGYVDSSTSPMKNTSAGN
VDIYRKIVHSDACSSKAVYSLRLACGAVLNSRSRSRIYGGSAALPGAMPQVSLHVO
NVHNGSGIITPEMIVTAACHYERKELNPMHMTARAGILROSEFYHGXQYQKISH
PNTDKTKNNDIALKLOKPLETFNLYKRVCLPNFGMMLQEPQDLWISMGWATERGK
TSEVLNAALVLIETORCSRVRVNLITTPAMICAGFLQGNVDSQGDSDGSLVTSNN
NIMWLIIGDTSWGSCAKAYRPGVYNVWFTDIWIVROMKANQ"
BASE COUNT 578 a 650 c 677 g 574 t
ORIGIN
Query Match 95.2%; Score 1655.2; DB 97; Length 2479;
Best Local Similarity 99.5%; Pred. No. 0;
Matches 1660; Conservative 0; Mismatches 8; Indels 0; Gaps 0;
```


us-09-615-285-1.rge

[illegible]

[illegible]

REFERENCE

id: 13596 GI: 6648959
 File: 13596.1
 house mouse
 Mus musculus
 Metazoa: Chordata: Craniata: Vertebrata: Euteleostomi:
 Eukaryota: Eutheria: Rodentia: Muridae: Murinae, Mus
 Mammalia
 1 (bases 1 to 153)

[illegible]

1583	AMCCCGAGCAGCATGTGGGTGGATTGGCTCCACAGAGACATGGCCCTGGCAGG	976
1584	TCAGCGTCGCGCGCGAGAGCTCAGCTGTGGAGGCTCCATCCATCACCCCGAGTGG	982
1585	TCAGCGTCGCGCGCGAGAGCTCAGCTGTGGAGGCTCCATCCATCACCCCGAGTGG	976
1586	TCAGCGTCGCGCGCGAGAGCTCAGCTGTGGAGGCTCCATCCATCACCCCGAGTGG	1036
1587	TCAGCGTCGCGCGCGAGAGCTCAGCTGTGGAGGCTCCATCCATCACCCCGAGTGG	1102
1588	TCAGCGTCGCGCGCGAGAGCTCAGCTGTGGAGGCTCCATCCATCACCCCGAGTGG	1096
1589	TCAGCGTCGCGCGCGAGAGCTCAGCTGTGGAGGCTCCATCCATCACCCCGAGTGG	1162
1590	TCAGCGTCGCGCGCGAGAGCTCAGCTGTGGAGGCTCCATCCATCACCCCGAGTGG	1156
1591	TCAGCGTCGCGCGCGAGAGCTCAGCTGTGGAGGCTCCATCCATCACCCCGAGTGG	1222
1592	TCAGCGTCGCGCGCGAGAGCTCAGCTGTGGAGGCTCCATCCATCACCCCGAGTGG	1216
1593	TCAGCGTCGCGCGCGAGAGCTCAGCTGTGGAGGCTCCATCCATCACCCCGAGTGG	1282
1594	TCAGCGTCGCGCGCGAGAGCTCAGCTGTGGAGGCTCCATCCATCACCCCGAGTGG	1276
1595	TCAGCGTCGCGCGCGAGAGCTCAGCTGTGGAGGCTCCATCCATCACCCCGAGTGG	1342
1596	TCAGCGTCGCGCGCGAGAGCTCAGCTGTGGAGGCTCCATCCATCACCCCGAGTGG	1336
1597	TCAGCGTCGCGCGCGAGAGCTCAGCTGTGGAGGCTCCATCCATCACCCCGAGTGG	1402
1598	TCAGCGTCGCGCGCGAGAGCTCAGCTGTGGAGGCTCCATCCATCACCCCGAGTGG	1396
1599	TCAGCGTCGCGCGCGAGAGCTCAGCTGTGGAGGCTCCATCCATCACCCCGAGTGG	1462
1600	TCAGCGTCGCGCGCGAGAGCTCAGCTGTGGAGGCTCCATCCATCACCCCGAGTGG	1456
1601	TCAGCGTCGCGCGCGAGAGCTCAGCTGTGGAGGCTCCATCCATCACCCCGAGTGG	1522
1602	TCAGCGTCGCGCGCGAGAGCTCAGCTGTGGAGGCTCCATCCATCACCCCGAGTGG	1516
1603	TCAGCGTCGCGCGCGAGAGCTCAGCTGTGGAGGCTCCATCCATCACCCCGAGTGG	1582
1604	TCAGCGTCGCGCGCGAGAGCTCAGCTGTGGAGGCTCCATCCATCACCCCGAGTGG	1576
1605	TCAGCGTCGCGCGCGAGAGCTCAGCTGTGGAGGCTCCATCCATCACCCCGAGTGG	1642
1606	TCAGCGTCGCGCGCGAGAGCTCAGCTGTGGAGGCTCCATCCATCACCCCGAGTGG	1636
1607	TCAGCGTCGCGCGCGAGAGCTCAGCTGTGGAGGCTCCATCCATCACCCCGAGTGG	1702
1608	TCAGCGTCGCGCGCGAGAGCTCAGCTGTGGAGGCTCCATCCATCACCCCGAGTGG	1696

AUTHORS	Jacqueline,E., Rao,N.V., Rao,G.V. and Hoidal,J.R.
TITLE	Cloning, genomic organization, chromosomal assignment and expression of a novel mosaic serine proteinase: epitheliasin
JOURNAL	FEBIS Lett. 468 (1), 93-100 (2000)
MEDLINE	20148617
REFERENCE	2 (bases 1 to 1753)
AUTHORS	Jacqueline,E.J., Rao,N.V., Rao,G.N. and Hoidal,J.R.
TITLE	Direct Submission
JOURNAL	Submitted (14-DEC-1998) Pulmonary Division, University of Utah, North Medical Drive, Salt Lake City, UT 84132, USA
FEATURES	Location/Qualifiers
SOURCE	1..1753
CDS	/organism="Mus musculus" /strain="BALB/c" /db_xref="taxon:10090" /chromosome="16" /map="telomeric region of long arm" 101..1573 /codon_start=1 /product="mosaic serine protease eptheliasin" /protein_id="AAFP1308.1" /db_xref="GI:6648960" /translation="MALNSGSPGICRGCEYENHGYOSEHCPEPPVPAVNGYNLYPAQY YPSPQVAPRITTOASTSVITHPRKSSGACITRSKRSLDLALATLTGTGAANVAW LIMRPWDNCSRTSEMECGLGTCTISSLMDCGVAHCPGEDENKVRILYGSFTLOY LSORAKYPCVDOMSENHYGRAACDKMYKNPFYSQSIPOQSATSPMKNLVSNGO DXYKVHSDSGSSRMVVSLRCIECGVSKROSRIYIGLNASPGDMOVSLVHVG HYCGSITTPPIVIAAHCVEEIPSLPMTAFAGILFOSLMFYGRHOYEKVISHP YDSKTRNDIDLMKLQTPLARLDIYKPCFLNPGMGLSDDOECRTISGMGAYEKGTST DVLANMAVPLEIPSCKNSRYIIYNNLIITPAKICAGTGLSDVSCQDGSGPLVTLKNG MWLIDTSWGSGCAALRPYGAVNTVETDWIYQMRANS"
BASE COUNT	461 A 438 C 455 G 399 T
ORIGIN	
Query Match	56.0% Score 972.6; DB 94; Length 1753;
Best Local Similarity	75.9%; Pred. No. 3.6e-248;
Matches 1262; Conservative	0; Mismatches 379; Indels 22; Gaps 4
Oy	92 atcgtgtgtaaaacgaagatggcttttgaaactcaagggtaccacaagctatacgacct 151 Db 81 ATGCTGCTGACACAGCGAGATGGCATTTGAACTCAGGGTGCCACTCCAGAAATCGAACCTT 140
Oy	152 actctgaaaacatagatccaacccgaaaaaaccccatctccgcagaagcccacgtgtgttc 211 Db 141 GCATTGAGAACCACGAGGATACAGTCTAGACACATCTGTCTCTCGAGACACACAGTGGCTC 200
Oy	212 ccactgtctacgaaggtagcatccggtcctaagtactaccgcgtgcccggtgcccaagtagcc 271 Db 201 CCATATGGCTACAATCTGATCACGCCACAGTACTACCATCTCCAGTGCCTCTAGTATGCTC 260
Oy	272 cgaagggtccctgaagcgaaggtctcaaacccccgcgtctctgcagcagaagcccaaatcccatccg 331 Db 261 CGAGGATTTACAACGGAACCTCTCAACATCTGTCTCATTCACACACACATGCCAAGTCC--TCAG 317
Oy	332 ggaacgtgtgcacactcaagaagactaagaagaagacacttgtcatacactcttgaaaccttgaggacct 391 Db 318 GAGCACCGGCGACCTCAAAAGTCTAGAATAATGCTGTGTTTAGCCCTTCCCTCGGGCACTG 377
Oy	392 tccctgttgaggagctgtgcgtctgcgcgtcttgacctactctgtgaagtctatgggcagaagtgct 451 Db 378 TCCTTACACGGGACCTCTGTGGCTGCTCTTCTTCTTGGAGGTTCGGGACAGCAACTGT 437
Oy	452 ccaactcttgagtagagtgcaactcctctcaagtagtacctgataaacaccttaactgtgtgt 511 Db 438 CTACGCTCAGATGAGGTGTGGGTCTCTAGGCACATGCAATGCAAGCTTCTCTCTGTGTGG 497
Oy	512 atggcggtgcaacactgtcccccgcgaggaggaagaagaatacgtgtgtcttcgcacctcagac 571 Db 498 ACCGGGTACACATTTGTCACACGGAGAAAGATGAGAACCTTGTGTGTCTCTCTACGGAC 557
Oy	572 caaaactcaactcagtgtaactactctcaagaagaaagtctggaaacctgtgtccaag 631

Db	558	AAACCTTCATCCTCCAGGTTTACTCATCTCCAGAGAAAGCTGTGTAATCCCGTGGCCAGG	617
Oy	652	acgactcggaaacgaactacagcgccgagcctcgcagggacacatggcctataagaataatc	691
Db	618	ATGATTTGGACGAGACAACTACGGGAGACACACATGTAAAGCAATGGGATACCAACAAATTT	677
Oy	692	ttactactagccaaaggaatagtgatgaacggggatccacacagcttataaactgaaca	751
Db	678	TTTTATTTCCAGCCAAAGGATTACCAACCAAGACGGGGCAACAGCTTTATACACTGAATG	737
Oy	752	caagtcgcggcaatgcgatatctataaaacacgttaaccaagtgatgcctgtctcaaa	811
Db	738	TCAGCTCAAGCCAAATTTBACCTCATATAAAAACTCAACACAGTACATGTTTCATCCC	797
Oy	812	aagcagtggtctcttlaacgtgtatagccctcgcgggtgtaacttgaaactcaagccagaa	871
Db	798	GCATGTGCTTTCTTTGGCTCTGTATAGAAATCCGGGGTTCGCT---CACTGAACGCCAGA	854
Oy	872	gcagagctcttgagcggcagagcgcgcctccgggggagcccttgacccttgacagctgc	931
Db	855	GCGAGATTGGTGGTGGATTGAATGGCTCACCAAGAGACTGGCCCCGCGACGCTGACCTGC	914
Oy	932	acgtccagaacgtlcaacgtctgcggaagctccataatacccccgagtgatctgtacag	991
Db	915	ACGTCCAAGGCGTCCACGCTGCGAGGCTCATATACCCCGAGTGGATTGTGACGG	974
Oy	992	cgcgcacactgcgtggaaaaaaccccttaacaccccaatgcatctggaacgcatctgcggga	1051
Db	975	CCGCCCACTGTGGAGAAACCCTCAGCGGGCCAGGACTGAGACGCATTTGGGGGAA	1034
Oy	1052	ttctgagacaatcttcaatgltctaltagagccgagataccaagtataaaaaagtgatcttc	1111
Db	1035	TTCTGAGACAGCTCTCATGTTCTATAGGAAGTAGACACCAGGTAGAAAAAGTAATTTCCC	1094
Oy	1112	atccaattatgctctcaaacaccaaaacaatgacatctgcctgtatgaagctgcagaaac	1171
Db	1095	ATCCAAATTCAGACTCTAAGACCAAGATATAGCAATTTGCTCTCATGAAAGCTGCAGACAC	1154
Oy	1172	ctctgactcttcaacgacactgataaacagtgctgtctgtcccaaccagcatgatgtctgc	1231
Db	1155	CTTTGGCTTTAATGATCTGATGTAAGCAAGCAAGTGTGTCTGCGCAACCAGGCAATGATGCTAG	1214
Oy	1232	agccagaaacgctctgtctgatatcttcgggtgtggggggccacgcgagggagaaaggaagacct	1291
Db	1215	ACCTAGACCAAGAAATGCTGGATTTCCGGGTGGGGGGCCACTCTATGAAAGGAAGAACCTT	1274
Oy	1292	cagaagctctgaacgcctgcgcaagtgctctcaltcaltgagacaagaagatgaacaagcaagt	1351
Db	1275	CGGACGTGTGAATGCTGCCATGTGTACCTTGATCGAGCCCTCCAAATGTAAATGTAAT	1334
Oy	1352	atgtctatgacaacctgtatctacacacagcatgatctgtccggctctcctcgaagggaagc	1411
Db	1335	ACATTTACAAACAACTTAATTCACACAGCATGATCTGTGCCGGCTTCTCCAGGGGCTG	1394
Oy	1412	tcgatctcttgcgaagtgacagctggaaggacctgtgtcaactcgaagaacaatatctggt	1471
Db	1395	TCGACCTTGGCCAGGAGACAGTGAAGGCGCGCTGTACTTTTAAAGATGGGATTCGGT	1454
Oy	1472	ggctgataagggaatacaagctggygttcttgctgtgcgcaaaagcttaacgacaaagatgt	1531
Db	1455	GGCTGTGTTGGGACACGACTTGGGCTCGGCTGTGGCCAAAGCACATCAGACTTGAGTAT	1514
Oy	1532	acgggaatctgaatgtatctcaacgagactgattatcgaacaaatgaagggcagacgctaat	1591
Db	1515	ACGGGAACGTGAAGGTATTTACAGATTGGATCTACACAGAAATGCGCGCAACGCTAAT	1574
Oy	1592	cccaatgg-----tctgctctgaacgtcglttacaagaanaaaatgg	1636
Db	1575	CCAAATGGCTTTGTCCAGACTTCCTTTGTTCTTCAACAACCTTTTGCAAGAAACCAAG	1634
Oy	1637	gctctg-gtttgtctcccggtgacatctactctatgaatgacatgacatgaaggtacatca	1695
Db	1635	GCGGAATTTAACTTCTCTGTGCACAATGATACCTTTTGATGATGATTCGAAGGCGCTTCA	1694

RESULT	7	1735 bp transmembrane P-
AF243500	AF243500	Plasmic
LOCUS	Mus musculus	
DEFINITION	AF243500.1	GI:7578918
ACCSSION		
VERSIONS		
REMARKS	Mus musc.	
SOURCE	Mus musculus	
ORGANISM	Eukaryota; Euteleostomi; Chordata; Vertebrata; Euteleostomi; Muraenae; Muraenae; Mus	

REFERENCE
AUTHORS Han, J. and Krasemembrane P-
TITLE putative
JOURNAL unpublished to 1735)
AUTHORS 2 (based Kim, S.
TITLE Han, J. Submission
JOURNAL Direct. Submission
10550 N. Torrey Pines Rd., La Jolla, CA 92126, USA
Local 735
musculus*

```

FEATURES
source
/organisms=
taxon:10030
xref="

```

[illegible]

	BASE COUNT	ORIGIN	DB 94:	Length 1000	Gaps
441 a	Score 969;	NO. 3.3e-247;	Indels 22;		

[illegible][illegible]

Db	1455	GGCGATTGGGAGACACAGCTGGGGCTCGGGCTGTGCCAAGCAGCAGCTACAGCTGGAGTAT	1514
Qy	1532	acgggaatgtgatgtatattcaacgactgtaattatcgacaatatgagygcaagcgcgtat	1591
Db	1515	ACGGGAACGTGACCGCTATTACAGATTGGATTCACAGCAAAATGAGGGGGAACAGCTAT	1574
Qy	1592	ccacatgt-----tcttcgcctctgcgcgtctgtttcaagaagaacatgtg	1636
Db	1575	CCACGTGGCTTGTGCCAGACTTCCTTTGCTTCAACAACTTCTCAAGAAAACCAAG	1634
Qy	1637	gagctg-gtttgcctccgcgtgcatagtattactcttaagaatcattcagagtcattca	1695
Db	1635	GCGCTAATTTAACTTCCTGTGCGACATGTATACCTTTGGATGATTCGAAAGGCGCTTCA	1694
Qy	1656	ctttattaaacagtgaactgtcttcgcaaaaaaaaaa	1736
Db	1695	CTTTTACTTAAACAGTGAAGCTGTTGACTGTGAAAAA	1735

FEATURES	SOURCE
LOCUS	AB038497
DEFINITION	AB038497 3507 bp mRNA for embryonic serine protease-2, complete cds.
ACCESSION	AB038497
VERSION	AB038497.1 GI:9757699
KEYWORDS	embryonic serine protease-2; Xenopus embryonic serine protease-2. Xenopus laevis cDNA to mRNA.
SOURCE	Xenopus laevis
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Amphibia; Batrachia; Anura; Mesobatrachia; Pipridae; Pipidae; Xenopodinae; Xenopus.
REFERENCE	1 (sites)
AUTHORS	Yamada,K., Takabatake,T. and Takeshima,K.
TITLE	Isolation and characterization of three novel serine protease genes from xenopus laevis
JOURNAL	Gene 252 (1-2), 209-216 (2000)
MEDLINE	20363741
REFERENCE	2 (bases 1 to 3507)
AUTHORS	Yamada,K., Takeshima,K. and Takabatake,T.
TITLE	Direct Submission
JOURNAL	Submitted (17-FEB-2000) to the DDBJ/EMBL/Genbank databases. Kazuo Yamada, Nagoya University, Graduate School of Human Informatics; Furo-cho, Chikusa-ku, Nagoya 464-8601, Japan (E-mail:yamadateinfo.human.nagoya-u.ac.jp, Tel:+81-52-789-2573)
FEATURES	Location/Qualifiers
SOURCE	1..3507

[illegible]

```

      .      ARANKGVYGNVTTFLEWISQMRITR
BASE COUNT      1035 a      712 c      740 g      1020 t
ORIGIN

```

Query Match	17.0%;	Score 295.6;	DB 8;	Length 3507;
Best Local Similarity	55.2%;	Pred. No. 8.3e-68;		
Matches 621; Conservative	0;	Mismatches 499;	Indels 6;	Gaps 2.

Oy	469	tgcaacatcccaagtaacccgacatcaaaccccttaacctggtgtagagagcggtccacatgc	226
Oy	469	tgcaacatcccaagtaacccgacatcaaaccccttaacctggtgtagagagcggtccacatgc	226
Oy	1694	tgtggcctacacagaaagttgtgctccgtcttccacagtggtgtagagggcgctgcacactgt	1753
Oy	529	ccggcgaggagagacggaatcggtgtgttgcccttaaggacccaacttcaactccctgac	588
Oy	1754	ccttatgtgagaaagatgaaatgctctgttgtagcttatatccacagacttccaaactgcag	1813
Oy	1814	gtctatttctacatcagtcagcagatggctgctgtgtgcagatgattactgcaaatgatgac	1873
Oy	649	tacggcgaggcgccctgcagggacatggtctataa---gaataatttactcttagccaa	705
Oy	1874	tttgggaaggtttggcctccacagacttttgatattaacggagcactatnacagatatgat	1933
Oy	706	ggaaatagtagtagcagcggaatccacccagattatgaactgaacaacatgctcgagcaat	765
Oy	1934	actttgatgtccctacatgacgcccttaatgggtattttcaagtttactctgatatcgagaa	1993
Oy	766	gtcgataltctaataaaaactgtaccacagtatgctgttcttcaagaagcagtggttct	825
Oy	1994	agcaaatctctatcacaaagtttcacatpacagtttactgattttctgaaacgttggtcttc	2053
Oy	826	ttacgtgtatagccttgagggttcaacttgaactccaagccgcagagacagattgctgagc	885
Oy	2054	tttactgtgataaagttgttggtgt---atctaatpacagtttactcagtgatgaatgtagga	2110
Oy	886	ggcgagagcgcgctcccgaggagccttgagccttgcaagtcagcgtcgacggtccagaagctc	945
Oy	2111	ggaaacttttggccaaacttggggagactggcagatcgccaaactcaatttaccatpacaggt	2170
Oy	946	caagtgctcgagaggtctcatatcatcaaccgcagatgagctgttgacaagccgcacatcgctg	1005
Oy	2171	gtttctatgtggtgggtccatctatttccaccacaaagtgcatagtactgcagctcactgtgta	2230
Oy	1006	gaaaaaccccttaacaatccatctgagatctgacgagcatltgcggagatlttgagaacatc	1065
Oy	2231	tatgggtcttactcaaaagtcactctgaaatggagagacttttctgtggagacactgaccaaacca	2290
Oy	1066	ttcaatctctatgtagccggaatcaacaaatagaaaaagagatctcatccaaatattgac	1125
Oy	2291	actttactatamgccagccctctattttctgtggaaagaaatattgttgacaccttggtattaac	2350
Oy	1126	ttcaagaaccaagaacaaatgacatctgcgctgtagaactctgcagaagacactctgactttcaac	1185
Oy	2351	ttctttacactttatgatatagcatatgcctcttaantgaamaactaacagatgaataaacttttggc	2410
Oy	1186	gaacttagtgaacacagtggtgtctgccaaaccacagcatgatagtctgcagccagaacagctc	1245
Oy	2411	ttataccacacagcggctgtctgttaccacaaactctggcagattttctggagagctggcacaca	2470
Oy	1246	tgcttgatctccgggtgagggggccacccggagggagaagaggaagacctctagaagtgctgaac	1305
Oy	2471	actttgagttatctgtttggggatctaaccttatgaaagagatagtgctccacatattctgcag	2530
Oy	1306	gctcgacaagtgctcttcatcttgacaacagagatgcaacacagatalgtctatgacaac	1365
Oy	2531	tatgcctgcattccacatgattgactgcgaatgtgtgcacatcattatgtgtacaaatgg	2590
Oy	1366	ctgatcacaccagcatgatactgtgctgcgtctccctgcaggggaaggttcgattcttgcag	1425
Oy	2591	caaatattacattctatagatatgtgctgttattcttcttctggggagatgatatacttggcag	2650

QY	1426	ggttcacagtggaggggcccctcgtgccttcacgaagaacaatctggtgtgcgtgaaggat	1485
Db	2651	GGGGACACGGGGGACCTTACTATTACAAAAGAAAATGGCATTGGTGGCTGTGTGTGAC	2710
QY	1486	acaagcttgggtcttcgagctgtgcacaaagcttcacagaccagagatgtacaggaatgtatg	1545
Db	2711	ACCACCTGGGGAGATGGTGTGTCGCCGAGCAATAAAGCTGGAGTATATGGGAATGTGACA	2770
QY	1546	gtattcacaagcagtgattatcgcacaatatgagggcagacagcgttaatt	1591
Db	2771	ACGTTCTTGGAATGATATATTACACAGATGAGAGCTTACAGATGAT	2816
RESULT	9		
LOCUS	AP001609/c		
DEFINITION	AP001609 Homo sapiens genomic DNA, chromosome 21, clone:KB447A5, MX1-D21S171	PRI	03-JUN-2000
ACCESSION	AP001609		
VERSION	AP001609.1	GI:7670563	
KEYWORDS	HTG.		
SOURCE	Homo sapiens pre-pro-B cell cell_line:FLB14-14 DNA, clone_lib:Keio		
ORGANISM	BAC library clone:KB447A5.		
REFERENCE	Homo sapiens		
AUTHORS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
TITLE	Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.		
JOURNAL	1 (bases 1 to 108927)		
AUTHORS	Shimizu,N., Kudoh,J. and Shibuya,K.		
TITLE	Homo sapiens genomic DNA, chromosome 21, clone:KB447A5, MX1-D21S171		
JOURNAL	region		
REFERENCE	Published Only in Database (2000) In press		
AUTHORS	2 (bases 1 to 108927)		
TITLE	Shimizu,N., Kudoh,J. and Shibuya,K.		
JOURNAL	Submitted (04-APR-2000) to the DDBJ/EMBL/GenBank databases.		
FEATURES	Direct Submission		
SOURCE	Location/Qualifiers		
	1. 108927		
	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
	/cell_line="FLB14-14"		
	/cell_type="pre-pro-B cell"		
	/chromosome="21"		
	/clone="KB447A5"		
	/clone_lib="Keio BAC library"		
BASE COUNT	29837 a 25350 c 26654 g 27086 t		
ORIGIN			
Query Match	12.8%; Score 223; DB 91; Length 108927;		
Best Local Similarity	100.0%; Pred. No. 1,8e-48;		
Matches	223; Conservative 0; Mismatches 0; Indels 0; Gaps 0;		
QY	127	gggttacaccacgactattggaacctactactatgaaaaacatgataaccacggaanaaccc	186
Db	97312	GGGTACACACACAGCTATTGTGACCTTACTATATAAACCATGATACCAACCGAAAAACCCC	97253
QY	187	tatcccgacagacgacactgtgtccacactctctacagagtgatcgcgcgcacagtaacac	246
Db	97252	TATCCCGACACAGCCACACTGTGTGTCCTCCACTGTCTTATCGAGAGGTGCATCCGGCTGAGACTAC	97193
QY	247	ccgtcccccgtgcacagtaacgcccgcgaggggttcctctgacagcagagcttcaaacccgctgc	306
Db	97192	CCGTCCCCCGTGGCCCACTAGACGCCCGGAGGGTCTCGACGACGAGCTTCAACCCCGTCTC	97133
QY	307	ttgcagcagcccaaatccccaatccgcggagacagtgtgcaactcaa	349
Db	97132	TGCACGACGCCCAATCCCAATCCGCGGACACTGTGCACCTTCA	97090

RESULT 10	AP001610	132110 bp	DNA	PRI	03-JUN-2000
LOCUS	AP001610/c				
DEFINITION	Homo sapiens genomic DNA, chromosome 21, clone: C172533B8,				
ACCESSION	MX1-D21S171 region, complete sequence.				
VERSION	AP001610				
KEYWORDS	AP001610.1	GI:7670564			
SOURCE	HTG.				
ORGANISM	Homo sapiens sperm DNA, clone_lib:C17-HSP BAC library clone:C172533B8.				
REFERENCE	Homo sapiens				
AUTHORS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.				
TITLE	1 (bases 1 to 132110)				
JOURNAL	Shimizu,N., Kudoh,J. and Shibuya,K.				
REFERENCE	MX1-D21S171 region				
AUTHORS	Published Only in Database (2000) In press				
TITLE	2 (bases 1 to 132110)				
JOURNAL	Shimizu,N., Kudoh,J. and Shibuya,K.				
REFERENCE	Submitted (04-Apr-2000) to the DDBJ/EMBL/GenBank databases.				
AUTHORS	Direct Submission				
TITLE	Nobuyoshi Shimizu, Keio University, School of Medicine, Molecular Biology, 35 Shinanomachi, Shinjuku-ku, Tokyo 160-8582, Japan				
JOURNAL	(E-mail:nshimizu@med.keio.ac.jp, Tel:81-3-3351-2370, Fax:81-3-3351-2370)				
FEATURES	Location/Qualifiers				
source	1..132110				
	/organism="Homo sapiens"				
	/db_xref="taxon:9606"				
	/cell_type="sperm"				
	/chromosome="21"				
	/clone="C172533B8"				
	/clone_lib="C17-HSP BAC library"				
BASE COUNT	36191 a 30522 c 32161 g 33236 t				
ORIGIN					
Query Match	12.8%; Score 223; DB 91; Length 132110;				
Best Local Similarity	100.0%; Pred. NO. 1.8e-48;				
Matches 223; Conservative	0; Mismatches 0; Indels 0; Gaps 0;				
OY 127	gggtacacacacagctatgagacctactatgaaacaacatgatgataccaacggaaaaacccc 186				
DB 61120	GGGTACACACACAGCTATTGGACCTTACTATGAAAAACAATGATACCAACCGGAAAAACCCC 61061				
OY 187	tatcccgacagacagccacgctgtgtccccaatgctcagagctgacacccgctgaactac 246				
DB 61060	TATCCCGACAGACACCCACATGTGTGCCCATGTCATCGAGAGTCAATCCGGCTCAGTACTAC 61001				
OY 247	ccgtcccccgtgtccccaatgacgaagcgcctgaacgagcttccaacccgctgcgc 306				
DB 61000	CCGTCCCCCGTGTCCCGACAGTACGCCCCCGAGAGGTCCTGACGACGAGCTTCCACACCCCGTCTGC 60941				
OY 307	tgcacgcagcccccaatccccatccgcggagacagtgcacactcaa 349				
DB 60940	TGCAcGcAGcCCCAATCCCAATCCCAATCCCGAGcAGcGTGcACACTCA 60898				
RESULT 11					
LOCUS	HS21C085	340000 bp	DNA	PRI	24-MAY-2000
DEFINITION	Homo sapiens chromosome 21 segment HS21C085.				
ACCESSION	AL163285 AP001740 BA000005				
VERSION	AL163285.2 GI:7717384				
KEYWORDS	human.				
SOURCE	Homo sapiens				
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.				
REFERENCE	1 (bases 1 to 340000)				
AUTHORS	Hattori,M., Fujiyama,A., Taylor,T.D., Watanabe,H., Yada,T., Park,H.S., Toyoda,A., Ishii,K., Totoki,Y., Choi,D.K., Soeda,E.,				

```

Ohki,M., Takagi,T., Sakaki,Y., Taudien,S., Blechschmidt,K.,
Polley,A., Menzel,U., Delabar,J., Kumpf,K., Lehmann,R.,
Patterson,D., Reichwald,K., Rump,A., Schillhabel,M., Schudy,A.,
Zimmermann,W., Rosenthal,A., Kudoh,J., Shibuya,K., Kawasaki,K.,
Asakawa,S., Shintani,A., Sasaki,T., Nagamine,K., Mitsuyama,S.,
Antonarakis,S.E., Minoshima,S., Shimizu,N., Nordstok,G.,
Hornischer,K., Brandt,P., Scharfe,M., Schoen,O., Desario,A.,
Reichelt,J., Kauer,G., Bloeker,H., Ramser,J., Beck,A., Klages,S.,
Hennig,S., Rlesselmann,L., Dagand,E., Wehrmeyer,S., Borzym,K.,
Gardiner,K., Nizetic,D., Francis,F., Lehnach,H., Reinhardt,R. and
Yaaso,M.L.
TITLE
JOURNAL
Submitted (03-MAY-2000) The Chromosome 21 Mapping and Sequencing
Consortium: * RIKEN Genomic Sciences Center, Human Genome Research
Group * Institute of Molecular Biotechnology, Genome Analysis *
Keio University School of Medicine, Dept. of Molecular Biology *
GBF, Dept. of Genome Analysis * Max-Planck Institute for Molecular
Genetics (addresses see below)
The Chromosome 21 Mapping and Sequencing Consortium consists of
* RIKEN Genomic Sciences Center, Human Genome Research Group, *
Sagamihara 228-8555, Japan,
* e.mail: sakaki@gscc.riken.go.jp
* URL: http://hgp.gsc.riken.go.jp/
and
* Institute of Molecular Biotechnology, Genome Analysis, *
Beutenbergstrasse 11, D-07745 Jena, Germany,
* e.mail: gscj-submit@genome.imb-jena.de
* URL: http://genome.imb-jena.de/
and
* Keio University School of Medicine, Dept. of Molecular Biology, *
Tokyo 160-8582, Japan,
* e.mail: shimizu@mb.med.keio.ac.jp
* URL: http://adenine.dmb.med.keio.ac.jp/
and
* GBF, Dept. of Genome Analysis,
* Mascheroder Weg 1, D-38124 Braunschweig, Germany, * e.mail:
info.genome@gbf.de
* URL: http://genome.gbf.de/
and
* Max-Planck Institute for Molecular Genetics,
* Inestrasse 73, D-14195 Berlin, Germany,
* e.mail: info-chr21@molgen.mpg.de
* URL: http://chr21.fz-berlin.mpg.de/
Location/Qualifiers
1. 340000
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="21"
/map="21q22.3"
<1. 81445
/note="Accession No. AJ011929"
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="21"
/map="21q22.3"
/clonename="P269A14, 5' partial"
/clonelib="RPC11.3-5 PAC library"
<1. 47014
/note="Accession No. AJ011929"
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="21"
/map="21q22.3"
/clonename="P141D16, 5' partial"
/clonelib="RPC11.3-5 PAC library"
3304. 3525
/note="MER58A"
/rpt_family="DNA/MER1_type"
/rpt_type=DISPERSED
4254. 4579
/note="AluSx"
/rpt_family="SINE/Alu"
/rpt_type=DISPERSED
FEATURES
source
1. 340000
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="21"
/map="21q22.3"
<1. 81445
/note="Accession No. AJ011929"
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="21"
/map="21q22.3"
/clonename="P269A14, 5' partial"
/clonelib="RPC11.3-5 PAC library"
<1. 47014
/note="Accession No. AJ011929"
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="21"
/map="21q22.3"
/clonename="P141D16, 5' partial"
/clonelib="RPC11.3-5 PAC library"
3304. 3525
/note="MER58A"
/rpt_family="DNA/MER1_type"
/rpt_type=DISPERSED
4254. 4579
/note="AluSx"
/rpt_family="SINE/Alu"
/rpt_type=DISPERSED
repeat_region
4693. 4768
/note="LINE/D"
/rpt_family="LINE/L1"
/rpt_type=DISPERSED
5206. 5229
/note="TTTTA)n"
/rpt_family="Simple_repeat"
/rpt_type=TANDEM
complement(5230. 5510)
/note="Aluub"
/rpt_family="SINE/Alu"
/rpt_type=DISPERSED
5518. 5541
/note="(T)n"
/rpt_family="Simple_repeat"
/rpt_type=TANDEM
5558. 5589
/note="Charlie7"
/rpt_family="DNA/MER1_type"
/rpt_type=DISPERSED
complement(5996. 6135)
/note="MIR"
/rpt_family="SINE/MIR"
/rpt_type=DISPERSED
complement(6240. 6541)
/note="AluSx"
/rpt_family="SINE/Alu"
/rpt_type=DISPERSED
7101. 7138
/note="(TTTA)n"
/rpt_family="Simple_repeat"
/rpt_type=TANDEM
complement(7139. 7415)
/note="L1PA5"
/rpt_family="LINE/L1"
/rpt_type=DISPERSED
complement(8130. 8232)
/note="L1PA5"
/rpt_family="LINE/L1"
/rpt_type=DISPERSED
8233. 8440
/note="L1PA5"
/rpt_family="LINE/L1"
/rpt_type=DISPERSED
8804. 9172
/note="MLT1B"
/rpt_family="LTR/MaLR"
/rpt_type=DISPERSED
11017. 11150
/note="MER5A"
/rpt_family="DNA/MER1_type"
/rpt_type=DISPERSED
complement(11278. 11746)
/note="L2"
/rpt_family="LINE/L2"
/rpt_type=DISPERSED
complement(11736. 12306)
/note="L2"
/rpt_family="LINE/L2"
/rpt_type=DISPERSED
complement(12307. 12591)
/note="AluSg"
/rpt_family="SINE/Alu"
/rpt_type=DISPERSED
complement(12592. 12783)
/note="L2"
/rpt_family="LINE/L2"
/rpt_type=DISPERSED
14096. 14164
/note="L2"
/rpt_family="LINE/L2"
/rpt_type=DISPERSED
14188. 14329

```

```

/note="LIMB3"
/rpt_family="LINE/L1"
complement(14330..14682)
repeat_region
/note="L1A1"
/rpt_family="L1R/MaLR"
/rpt_type="DISPERSED"
14685..15070
/note="LIMB3"
/rpt_family="LINE/L1"
/rpt_type="DISPERSED"
complement(17893..18027)
repeat_region
/note="MIR"
/rpt_family="SINE/MIR"
/rpt_type="DISPERSED"
18438..19329
/note="MER49"
/rpt_family="L1R/MER4-group"
/rpt_type="DISPERSED"
complement(20328..20688)
repeat_region
/note="MT1A1"
/rpt_family="L1R/MaLR"
/rpt_type="DISPERSED"
complement(20741..20997)
repeat_region
/note="L1R16C"
/rpt_family="L1R/Retroviral"
/rpt_type="DISPERSED"
21744..22035
repeat_region
/note="Alu5g"
/rpt_family="SINE/Alu"
/rpt_type="DISPERSED"
complement(22282..22412)
repeat_region
/note="MIR"
/rpt_family="SINE/MIR"
/rpt_type="DISPERSED"
complement(22950..23031)
repeat_region
/note="L2"
/rpt_family="LINE/L2"
/rpt_type="DISPERSED"
complement(25177..25479)
repeat_region
/note="Alu5g"
/rpt_family="SINE/Alu"

```

Query Match 12.8%; Score 223; DB 92; Length 340000;
 Best Local Similarity 100.0%; Pred. No. 1.8e-48;
 Matches 223; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 127 gggcaccacacagctatggacctaataaaccatgataccacggaaccc 186
 |||||
 DB 303106 GGGTCACACACACTTGTGACCTTACTATGAAACCATGATACCAACCGAAGAACCC 303047

QY 187 tatccgacacagcactgtgtcccaactgttaagaggtgataccggtcagtactac 246
 |||||
 DB 303046 TATCCCGCAGCCACTGTGTCCCACTGTCTACGAGGTGATCCGGCTCAGTACTAC 302987

QY 247 ccgtcccccgtgcccagtaagcccgaggtctctagcagagctcccaacccgctcgtc 306
 |||||
 DB 302986 CGGTCCCGGTGCGCCAGTACGCGCCGAGGCTCTGACGAGGCTTCCACCCCGTCTGC 302927

QY 307 tgcagcagacccaatcccatccggagagtgtagacctcaa 349
 |||||
 DB 302926 TGCACGACGACCCCAATCCCATCCGCGAGACAGTGTGACCTCA 302884

RESULT 12
 AB038159 2135 bp mRNA PRI 17-JAN-2001
 LOCUS AB038159 2135 bp mRNA PRI 17-JAN-2001
 DEFINITION Homo sapiens TMPRSS3c mRNA for serine protease, complete cds.
 ACCESSION AB038159
 VERSION AB038159.1 GI:12246827
 KEYWORDS serine protease.
 SOURCE Homo sapiens liver and thymus cDNA to mRNA.
 ORGANISM Homo sapiens

```

REFERENCE
AUTHORS
1 (sites)
Scott,H.S., Kudoh,J., Wattenhofer,M., Shibuya,K., Berry,A.,
Chrast,R., Guipponi,M., Wang,J., Kawasaki,K., Asakawa,S.,
Minoshima,S., Yonush,F., Mehdi,S.O., Radhakrishna,U.,
Papavasvas,M.P., Gehrig,C., Rossier,C., Korostilshesky,M., Gal,A.,
Shimizu,N., Bonne-Tamir,B. and Antonarakis,S.E.
Insertion of beta-satellite repeats identifies a transmembrane
protease causing both congenital and childhood onset autosomal
recessive deafness
Nat. Genet. 27 (1), 59-63 (2001)
JOURNAL
MEDLINE
20578749
REFERENCE
AUTHORS
Shimizu,N., Kudoh,J. and Shibuya,K.
Direct Submission
Submitted (10-FEB-2000) Nobuyoshi Shimizu, Keio University, School
of Medicine, Molecular Biology; 35 Shinanomachi, Shinjuku-ku, Tokyo
160-8582, Japan (E-mail:nshimizu@med.keio.ac.jp,
tel:81-3-3351-2370, fax:81-3-3351-2370)
FEATURES
Source
1..2135
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="21"
/map="21q22.3"
/tissue_type="Liver and thymus"
250..1233
/gene="TMPRSS3c"
250..1233
/gene="TMPRSS3c"
/note="TMPRSS3c"
/codon_start=1
/product="serine protease"
/protein_id="BAB20079.1"
/db_xref="GI:12246828"
/translation="MCSDDMKGHYANVACAOIGFPEYSSNDLRVSLRGPREPVS
IDHLPPDKVTALHSHSVYREGCAASHVYCTICAGHRGYSRRVGNMILLQMP
WQNSLOPQYHLCGGSVITPLMTTAAHCVYDYLKPSFTIOYGLVSLDNPAPSHV
EKIVYHSKTPKRLGNDIALMKIAGPLTNEMLQPVLSNSENPEPDGVVWNTSGWA
TEBGADGASPVLNHAAPLISNKCINHRDVGIIISPMICAGYLTGVDSCQGDSSG
PLVCOERRMLKLVGATFSGICAEVKKPVYTRVTSFLIMIHQMRDLKT"
BASE COUNT 487 a 582 c 504 t
ORIGIN
```

Query Match 12.2%; Score 212.2; DB 85; Length 2135;
 Best Local Similarity 54.2%; Pred. No. 1.4e-45;
 Matches 624; Conservative 0; Mismatches 488; Indels 39; Gaps 8;

QY 469 tgcgactccctcaggtactatcatcaacccttaactggtgtagcgtgtacactgc 528
 |||||
 DB 103 TGTGCTCATCTTTAAGTATATGACGTAGCTGATGATGACGAGTCTCGATTGC 162

QY 529 cccgacgaggaagacgaatcggtgttcgccttaagacaaactcatccttcag 588
 |||||
 DB 163 AAGAGCGGAGGAGGACAGATCCCGCTGTGCGGTGGGTGTGATAGATGCCGTCTCAG 222

QY 589 gtgtactcatctcagaggaagctcgtgacacctgtgtgccaagaacagcaggaac 648
 |||||
 DB 223 GTGTTTCACACTCTCT-----TCGTGGAAGACCATGTCCTCGATGCTGAAGGGTCA 276

QY 649 taagggcgggcgccctgcgaagagacatgaggtataagaataatttacttaaccaga 708
 |||||
 DB 277 TACGCAAAATGTCTCTGTGCCCAACTGCGTTTCCCAAGCTATGTGAGTTAGATACTC 336

QY 709 atagtg-gaagacagcagacacacagcttaagaaactgaacaagtcgagcaatgt 767
 |||||
 DB 337 AGAGTAGCTCGCTGAGGCGGAGATTCCGGAGAGAGATTGTGCCATGATACACTCTTG 396

QY 768 cgatactataaaaaactgtaccac-----agtatgcctgttcttca 810
 |||||
 DB 397 CCAGATGACAAAGTGACTGATACACACTCACTATATGTATGTAGAGGAGGATGTGCTCT 456


```

Db 790 GGCACAGTGTACCTTCAGTGCACAGCCCTGTGTCATAGAAAGGGCTACAGCTCAC-- 847
Oy 871 agcagagatgttgccgagagcgagcgctccgggggcttgccctggagagatcagccg 930
Db 848 ----GCATGCTGGGTGGAACAATGCTCTTCTCTGCACTGGCCGGGCGACCTT 903
Oy 931 cagctcagaacgacgctgtctgagaggtccatcatcaaccccgagatgagctgagaa 990
Db 904 CAGTTCAGAGGCTACACCTGTGGGGGCTGTGTCATCAGCCCTGTGATCATCAGT 963
Oy 991 gccgccaactgctgagaaacaccttaacaatcagatgagatgagcagatctgcggg 1050
Db 964 GCTGCACATGCTTATATA--CTTGTACCTCCCGCAAGTCATAGA---CCATTCAGGTG 1018
Oy 1051 attttagaagaatcttctatgtctatagagccgagatgaccagatgaaaaagatctct 1110
Db 1019 GCTGATTTCCCTGTTGGCAATCCA--GCCCATTCCTGTTGTTGAGAGATTTCTTAC 1077
Oy 1111 catcaaatatgactcacaacgaacaaatgacattgctgtatgagctgcagaaag 1170
Db 1078 CACAGCAATGACAGCAAGAGAGGCTGGGCAATGACATGCGCTTATGAAAGTGGCGGG 1137
Oy 1171 cctctgacttcaagagcactagtgaaacagatgtctgcgccaaccaggaatgagctg 1230
Db 1138 CCATTCACCTTCATGTAATGATCCAGCCTGTGCTGCTGCCCAACTCTGAAAGAACTTC 1197
Oy 1231 cagcagaacagcgtctgtgagatcttcgggtggtgggagccagcagaaga---gaaaggag 1287
Db 1198 CCGCATGCAAAAGTGTGCTGAGACGTAGAGATGGGGGCGACAGAGATGAGCAGGTGAC 1257
Oy 1288 accctagaagatgtcgaacgctgcgaaggtgtctctcatatgagacacagatgcagcag 1347
Db 1258 GCGTCCCGTCTGTGAAACAGCGGGCGGCTCCCTTTGATTTCACAAAGATCTGCACAC 1317
Oy 1348 agatatgtctatgacaacactgatcacacagcagatgctgtgcggtctctgtcaggg 1407
Db 1318 AGGAGAGTGTAGGTGAGTGCATCTCCCTCATGCTCTGCGCGGGGACCTGACGTGAGGT 1377
Oy 1408 aacgtgactcttcgacaggtgacagtgagtgagggcctctgttcaacttcgagaaacatlc 1467
Db 1378 GCGGTGACAGCTGCCAGGGGAGACCGGGGGGCGCTGTGTGTAAAGAGAGAGCTG 1437
Oy 1468 tgggtgctatagagggatcaacgctggtgtctgtgctgtgcgaacgcttcaagaccag 1527
Db 1438 TCGAAGTTAGTGGGAGCGACAGCTTTGGCATCGGCTGCGAGAGGTGAACAAGCTGG 1497
Oy 1528 gtgtacggagatgagtgatattcaacgagctgagatattatcgacaatgagggcagcgc 1587
Db 1498 GTGTACACCCGTTGTCACCTCTTCTGTGAGTGTGATCCAGAGAGATGGAGAGACCTA 1557
Oy 1588 tatccacatg 1598
Db 1558 AAAACCTGAAG 1568

```

RESULT 14
AB038158
LOCUS AB038158 2554 bp mRNA PRI 17-JAN-2001
DEFINITION Homo sapiens TMPRSS3b mRNA for serine protease, complete cds.
ACCESSION AB038158.1 GI:12246825
VERSION AB038158.1 GI:12246825
KEYWORDS serine protease.
SOURCE Homo sapiens liver and thymus cDNA to mRNA.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
AUTHORS Scott,H.S., Kudoh,J., Waltenhofer,M., Shibuya,K., Berry,A.,
 Chaoist,R., Guiponi,M., Wang,J., Kawasaki,K., Asakawa,S.,
 Minoshima,S., Yonush,F., Mehdi,S.O., Radhakrishna,U.,
 Papasavvas,M.P., Gehrig,C., Rossler,C., Korostilshesky,M., Gal,A.,
 Shimizu,N., Bonne-Tamir,B. and Antonarakis,S.E.
TITLE Insertion of beta-satellite repeats identifies a transmembrane

protease causing both congenital and childhood onset autosomal
 recessive deafness
 Nat. Genet. 27 (1), 59-63 (2001)
JOURNAL 20578749
MEDLINE 2 (bases 1 to 2554)
REFERENCE Shimizu,N., Kudoh,J. and Shibuya,K.
AUTHORS Shimizu,N., Kudoh,J. and Shibuya,K.
TITLE Direct Submission
JOURNAL Submitted (10-FEB-2000) Nobuyoshi Shimizu, Keio University, School
 of Medicine, Molecular Biology; 35 Shinanomachi, Shinjuku-ku, Tokyo
 160-8582, Japan (E-mail:shimizu@med.keio.ac.jp,
 Tel:81-3-3351-2370, Fax:81-3-3351-2370)
FEATURES
 source
 1..2554
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="21"
 /map="21q22.3"
 /issue_type="Liver and thymus"
 /gene="TMPRSS3b"
 /transcript="TMPRSS3b"
 /note="transmembrane protease, serine 3"
 /product="serine protease"
 /protein_id="BAB20078.1"
 /db_xref="GI:12246825"
 /translation="MCSDDMKGHVYANVACQQLGFPYSYSSDNLRLVSSLEGQREPFVS
 IDHLIPDDVYTAHLHSVYVREGASGHHVTLCTACGHRGYSSTIVGNNLSLQMP
 MQASLPQFQYHLGSGSVLPWIIIAHVCVYDLYLPKSWTIOVGLVSLIDNPAPSHLV
 EKIVYHSKYPKRLNDIALMKLAGLTFENKIOPLSPNSPENPDGVCWTSQMG
 TEDGADSPKRVNHAAPLISMKICNHRPVGIIISPSMLCAGYLTGGVDSQSGSG
 PLVCCERMLKVGATRSRGICAPVKNREVYRVSFLDIHDEMRDLKT"
 BASE COUNT 590 a 679 c 662 g 623 t

Query Match 12.2%; Score 212.2; DB 85; Length 2554;
 Best Local Similarity 54.2%; Pred. No. 1.4e-45;
 Matches 624; Conservative 0; Mismatches 488; Indels 39; Gaps 8;

```

Oy 469 tgcgactcctcgaagctgacatgacacccctcaactggtgtgagtgagctgacacatgc 528
Db 522 TGTGCTCATCCCTTAAGTGTATGAGCTGATAGCTGATGATGACGAGTCTCGAATTGG 581
Oy 523 cccgaggggagagagagaaatcggtgtgtgtgccttaagaccacaaatcactcctcag 588
Db 582 AAAGACGGGGAGAGACGAGTACCCCTGTCTCCGGGTGGTGTGAGATGCGCTGCCAG 641
Oy 589 gtgtactatctcagaggaagctcggccacctgtgtgccaagaagcagatggaacgaagc 648
Db 642 GTGTTACAGACTGCT-----TGTGGAAGACCATGTGTGCTCGATGACTGGAAGGTGAC 695
Oy 649 tacggcgcgcgctgcagagagacatgagcgtataaataatatttcaacttaagccaaga 708
Db 696 TACGCAAAATGTTGCTGTGGCCAACTGGGTTTCCAAAGCATGTGATGATCAGATACTCC 755
Oy 709 atagtg-gatgacagcggatcccaacgcttatgaaactgaaacagaatggtcggaagt 767
Db 756 AGAGTAGGCTCCGTGGAGGGGCGAGTTCGCGGAGAGCTTGTGTCATCATCACTCTTG 815
Oy 768 cgaatcataaaaaacttaccac-----agtgatcctgttcttca 810
Db 816 CCAGATGACAGGTGACTGATTCATACACACTCACTGATATGTGAGGAGGAGATGTGCTCT 875
Oy 811 aagcgagtggtcttcaactgactgataagctgcgggggtcaactgtaactgaacgcgcag 870
Db 876 GGCACAGTGTGTTACCTTGGCAGTGCACAGCCTGTGTCATAGAAAGGGGCTACACCTAC-- 933
Oy 871 agcagagatgtggggcgagagagcgctccgggggctgtgacctgcagtgacagctg 930
Db 934 ----GCATGTGGTGGTGAACAATGCTCTTCTCTCCAGTGGCGCTGGCAGGCGACCTT 989

```

OY	931	cgagcgcgaagaagcccaagctgttgcggagagctcccaatccaccccgagtgatcgttgaca	990
Db	990	CAgTTCaAGGGGTACCAcCTGTGCGGGGGCTGTGATCAACAGCCCTGTGGATATATACCT	1049
OY	991	gcgcgcacatgcgtgttgaaaaaaacctcttaacatccatgatgcatttgcagcgcatttgcggg	1050
Db	1050	GCTGCACACCTGTTTATGCA--CTTGTACTCTCCCAAGTATGGA---CCATCCAGGTGG	1104
OY	1051	attttgagagcaactcttcaattgtctatayagccggatccaagttagaaaaagtatcttct	1110
Db	1105	GTCTAGTTTCCCTGTGGACAATCCA-CCCCATTCCCAATTGTGGAGAAAGATTGTCTAC	1163
OY	1111	catccaattatgaacttccaaagcaagaacatgaactttgcctgcgtgaagcttcagaaag	1170
Db	1164	CACACCAAGTACAGCCAAAGAGCTGGGCAATATACATCGCCCTTATATGAAGCTGGCCGG	1223
OY	1171	cccttgactcttaacagcactagtgaaacagtgctgtctgcgcacccagatgatgtctg	1230
Db	1224	CCACTCACGTTTCAATGAAGATATATCCAGCCTGTGTGCCCTGCCCACTGTGAAGACAACTTC	1283
OY	1231	cagccagaagacgctctgctgattccgggtgggggggcccacagaga---gaaagggag	1287
Db	1284	CCCCATGGAAAAAGTGTGCTGGACGTCAAGATGGGGGCCACAGAGATGGACAGAGTGAC	1343
OY	1288	aacctcaagaagtgccttgaacgctctgcacaaagtgctctcatttgcattgagacacagatgcagaacg	1347
Db	1344	GCCTCCCTGTCCTGAACACAGCGGGCGCTCCCTTTGATTTCACAAAGATCTGCACAC	1403
OY	1348	agatatgtctattgaacaaactgataccaacagcaatgatctgtgcggctctccctgcagagg	1407
Db	1404	AGGCACTGTACGGGTGGGATCATCTCCCTTCATAGCTCTGCGGGCTACCTTCACCGGT	1463
OY	1408	aacgcgcgaattcttcgcagagctgcacagtgaagggcctcgtgcactcttgcgaagaacaaatc	1467
Db	1464	GGCGTGGACAGCTGCCAGAGGGGACAGCGGGGGCCCTGGTGTCAAGAGAGAGCTG	1523
OY	1468	tgtgtgcgtgaagggatatacaagcttgggtgttcgtgcgtgtgcacaaagcttacagacaga	1527
Db	1524	TGGAGTTTAGTGGGAGCACACGATTTGGCATTCGGCTGCCGAGAGGTAAACAAACCTGGG	1583
OY	1528	gtgtacgggaatgtatgtatgtatctcaagacatggatcttaacgaacatgaagggcagagcgc	1587
Db	1584	GTGTACACCCTGTACTCTTCTGTGACATGCCAGAGCAGATGACAGAGACTTA	1643
OY	1588	taatccacatcg	1598
Db	1644	AAAACCTGAAG	1654
RESULT 15			
LOCUS	AF201380	2416 bp	mrna PRI 03-DEC-2000
DEFINITION	Homo sapiens serine protease TADG12 mRNA, complete cds.		
ACCESSION	AF201380		
VERSION	AF201380.1	GI:11527821	
KEYWORDS	.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
AUTHORS	1 (bases 1 to 2416)		
TITLE	Underwood, L.J., Shigemasa, K., Tanimoto, H., Beard, J.B., Schneider, E.N., Wang, Y., Parmley, T.H. and O'Brien, J.T. Overlapi tumor cells express a novel multi-domain cell surface serine protease		
JOURNAL	Biochim. Biophys. Acta 1502 (3), 337-350 (2000)		
PUBMED	11068177		
REFERENCE	2 (bases 1 to 2416)		
AUTHORS	Underwood, L.J., Tanimoto, H., Schneider, E., Wang, Y., Parmley, T., Shigemasa, K. and O'Brien, J.T.		
TITLE	Direct Submission		
JOURNAL	Submitted (02-NOV-1999) OBGYN, University of Arkansas for Medical Sciences, 4301 W. Markham St., Little Rock, AR 72205, USA		

FEATURES	Location/Qualifiers
Source	1..2416 /organism="Homo sapiens" /db_xref="taxon:9606"
CDS	144..1511 /codon_start=1 /product="serine protease TADG12" /protein_id="AAG37012.1" /translation="MGENDPPAVEAPSFRLFGIDLDKISPAVDADAQAQLSLLE PFEFVSQSSSLGIIALIALAIGLIDHFCDSGKTRCRSSFCIDLIRCDVSDCKDQ EDVEYCAVVGOMNALQVFTAAWMKTMDSMDMKHAYANACDGFYSVSDMLRV SLEGEFRFEFVSIIDLIPDDKYTLAHSVYREGASHHVYTLCCGCHGRGYSRIR VGNMSLISQMPQASIOFOGYHLGGSVLPIVITRAHVVDLYLIPKSTTOGVIV SLDNPAPSHLYEKIVYHSKTRPKRLGNDIALMLAGLPLTNEIOPICUPENSENR DGKATWISGKATDEGDSPVLAHNAAPVLAISNKDLOQGVNRHHVLAHLRLPLQG WRMNSCQDGSGLVCQDERRLMKLVGATSFEGICADVNRPGVYTRVTSFLDWIHEQME RDLYR"
BASE COUNT	555 a 639 c 636 g 586 t
ORIGIN	
Query Match	12.0%: Score 208.6; DB 89; Length 2416;
Best Local Similarity	54.6%: Pred. No. 1.2e-44;
Matches 628; Conservative	0; Mismatches 484; Indels 39; Gaps 9;
469	tcgcagctctcaggtacccgcatccatcaaccctctaaactggtgtgatgagcgtgcacacgtc 528
381	tctccgtctatccctttrtaagctgtrattccagctgatatcactcagatgtgacggatctcggatttgc 440
529	cccgagcgaggagacgagatcgtgtgtgttcgctcttaacgaccaaacttcactcctaag 588
441	aaagacggggagagacgagatccgtgtgtccgggtgggggtgcacaaatgccgtctccag 500
589	gtgtactcaatccagagagatccctggacccctgtgtgacgaagagactggaagagagac 648
501	gtgtttcacagctgct-----tctgtggaagacatattgctccgattgactggagaggtcac 554
649	tacggcgaggcgccctgcagcgagacatgtggactaagaataatttctaacttgcacaaaga 708
555	tacgcaaatgttgccctgtgccacacgtgggtttcccaagctatgtgagttacagatmaaccgc 614
709	atagtg-gatgacagcgatccacacgcttcttgaaactggaacaaagtgcgcggaatgt 767
615	aagatgagcttccctcgaggggacagttccgggagaggttgttcacatgattaccctcttg 674
768	cgatatctctaaaaaactgtaccac-----agtgatgcctgttcttcca 810
675	ccagatgacaaagtgactgcatttacaaccacactagatattgtgagggagagatgtgcctct 734
811	aaagcagtggttcttcttaagctgtatagcctgcgcgggttcaacttgaactcaagccgcag 870
735	ggccacgtgtgttacctgtgcacagcacacgcttgtgtcatagaaaggcggtacagctcac-- 792
871	agcaagattgtgagcgcgagagcgcgctcccgaggggccgtgccccttgacagtgacgctg 930
793	-----gcattgtgtgggtgaamaactgttccttgcctgtgcacatggcctctggcagccacgctt 848
931	caagtcacaaagtcacagtgctgcgagaggtccatcatcaatcccccagatgtgactgtgaca 990
849	cagttccacagggctacacacctgtgcggggcctgtgtcatatcacccctctgtgattcattc 908
991	ggcgcccaactgtgtygaaaaaactctttaaacttcatacgatgtgaagcatttcgag 1050
909	gctgcacactgtgttatca--cttgtactctcccaagtcacagga--ccatccacaggtgg 963
1051	attttgagcaactcttctatgttctatgtaggcggatcccaagtgtaaaagttaattct 1110
964	gcttattgttccctgttgagcatatca--gcccacattccacacttggtgagagaagtgttttac 1022
1111	catcaaatatgactccaagaaccaagaacatgacattgcgtgtgataagctgcagaag 1170
1023	cacaccagatcacacccaagaaggcgtggcgcatgtacatgcctcttattaaagctggccggg 1082

THIS PAGE BLANK (USPTO)

RESULT	1	ALIGNMENTS
AAZ29636	AAZ29636 standard; cDNA, 1738 BP.	
XX AC	AAZ29636;	
XX DT	22-MAR-2000 (first entry)	
XX DE	Human 20PF12-GTC2 encoding cDNA.	
XX KW	20PF12; TMRRS2; androgen; serine protease; 20PF12-GTC1; cancer;	
XX KW	transmembrane protein; colon; prostate; prostate tumour; ss.	
XX OS	Homo sapiens.	
XX FT	Key	Location/Qualifiers
XX FT	CDS	112..1590
XX FT		/ftag= a
XX FT		/product= "20PF12-GTC2"
XX FT		/note= "Prostate-specific serine protease"
PN	W0962942-A2.	
XX PD	09-DEC-1999.	
XX PE	01-JUN-1999.	99WO-US12253.
XX PR	01-JUN-1998.	98US-0087598.
XX PR	29-JUN-1998.	98US-0091474.
XX PR	14-APR-1999.	99US-0129521.
XX PA	(UROC-) UROGENESYS INC.	
XX PA	(AFAR/) AFAR D E.	
XX PA	(HUBE/) HUBERT R S.	

PA (LEON/) LEONG K.
 PA (BAIT/) RAITANO A B.
 PA (SAFF/) SAFFRAN D C.
 PI Afar DE, Hubert RS, Leong K, Raitano AB, Safran DC;
 XX MPI: 2000-116363/10.
 DR P-PSDB: AMV44406.
 XX
 PT Novel cell surface antigen useful to treat colon and prostate cancer -
 PS
 XX Claim 2: Fig 1: 58bp: English.
 CC The present sequence is a cDNA encoding 20P1F12 protein (also known as
 CC the TMPRSS2 protein), which is a prostate-specific, androgen-regulated,
 CC cell surface serine protease. It is a glycosylated type II transmembrane
 CC protein with an extracellular C-terminal protease domain. Host cells can
 CC be transformed to produce the protein, using vector containing
 CC 20P1F12/TMPRSS2 gene (also designated 20P1F12-CTC1, as deposited with
 CC ATCC accession number 207097). Anti-20P1F12/TMPRSS2 antibodies may be
 CC used as therapeutic agent for prostate and colon cancers, to image
 CC prostate cancer cells and prostate tumours, to identify ligands and
 CC cellular constituents that bind to a 20P1F12/TMPRSS2 gene product and for
 CC use as cancer vaccines.
 CC
 XX Sequence 1738 BP: 433 A; 459 C; 473 G; 373 T; 0 other:

Query Match 100.0%; Score 1738; DB 21; Length 1738;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1738; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ggcggagggcggggagggcggaagggcgggcgccgctcgagcgcgaggtcat 60
 Db 1 ggcggagggcggggagggcggaagggcgggcgccgctcgagcgcgaggtcat 60
 QY 61 atctgaacatctcagatatacatatcactcgtatctgtctgtataacagcaagatggttg 120
 Db 61 atctgaacatctcagatatacatatcactcgtatctgtctgtataacagcaagatggttg 120
 QY 121 aactcaggggtccacacacacatctatgtacattactatgaataacatgataacacgggaa 180
 Db 121 aactcaggggtccacacacacatctatgtacattactatgaataacatgataacacgggaa 180
 QY 181 aaccctatcccgacagcgacactgtgtgtccacactgtctacagaggtgcatccgctcag 240
 Db 181 aaccctatcccgacagcgacactgtgtgtccacactgtctacagaggtgcatccgctcag 240
 QY 241 tactaccggtcccggtgcccagtaagcccgagggtccctgtagcagagcttccaaacc 300
 Db 241 tactaccggtcccggtgcccagtaagcccgagggtccctgtagcagagcttccaaacc 300
 QY 301 gtccgtctgacgagacgcaaatcccatcccgagagctgtgcaacctcaagactaagaa 360
 Db 301 gtccgtctgacgagacgcaaatcccatcccgagagctgtgcaacctcaagactaagaa 360
 QY 361 gcaactgtgatacacccttgaccccttgaggagacttccctcgttgaggagctgagcgctgagc 420
 Db 361 gcaactgtgatacacccttgaccccttgaggagacttccctcgttgaggagctgagcgctgagc 420
 QY 421 ctactcttgagaagtctcatgtggcagaaagtgtctccaactctgggataagtgcaattccta 480
 Db 421 ctactcttgagaagtctcatgtggcagaaagtgtctccaactctgggataagtgcaattccta 480
 QY 481 ggtacacgtgacaaacccttaactgtgtgatagtgcatgtgcacacgtgccccgagggag 540
 Db 481 ggtacacgtgacaaacccttaactgtgtgatagtgcatgtgcacacgtgccccgagggag 540
 QY 541 gaagagaaatcgtgtgtctgcaccttaagaaacaaacttcaatcccttaagtgatcatct 600
 Db 541 gaagagaaatcgtgtgtctgcaccttaagaaacaaacttcaatcccttaagtgatcatct 600
 QY 601 cagaggaagtcctctgacacctgtgtgcacaaagacgtggaacgagaacatacgggagcg 660

Db 601 cagaggaagtcctctgacacctgtgtgcacaaagacgtggaacgagaactacgggagcg 660
 QY 661 gctctgagagagacatgggtatagaataatcttactcttaagcagaagaaatgtgagac 720
 Db 661 gctctgagagagacatgggtatagaataatcttactcttaagcagaagaaatgtgagac 720
 QY 721 agcgatccaccagcttatgtaaactgaaacaagctgcccgaatgtctgatalatanaa 780
 Db 721 agcgatccaccagcttatgtaaactgaaacaagctgcccgaatgtctgatalatanaa 780
 QY 781 aaactgtacacagtgatgctgttcttcaaaagcagtggttcttcaactgtatagcc 840
 Db 781 aaactgtacacagtgatgctgttcttcaaaagcagtggttcttcaactgtatagcc 840
 QY 841 tgggggtgcaacttgaaacttaagcccgacagagcagatgtgtggcgaggaagcgctc 900
 Db 841 tgggggtgcaacttgaaacttaagcccgacagagcagatgtgtggcgaggaagcgctc 900
 QY 901 ccgggggcttgcccttgacaggtcagccttgacagctccagaacgttcacagtgtcggaagc 960
 Db 901 ccgggggcttgcccttgacaggtcagccttgacagctccagaacgttcacagtgtcggaagc 960
 QY 961 tccatcatcaccgcccgagtgatcgtgacacgcccacctgctgtggaaaaaacctctaac 1020
 Db 961 tccatcatcaccgcccgagtgatcgtgacacgcccacctgctgtggaaaaaacctctaac 1020
 QY 1021 aatcatgtgcatctgacagcatttgcgggatttgaagaacaacttcatgtatgta 1080
 Db 1021 aatcatgtgcatctgacagcatttgcgggatttgaagaacaacttcatgtatgta 1080
 QY 1081 gccggtataccaaagtgaataaagtgtattccatccaaatltatgactccaagaacaaagac 1140
 Db 1081 gccggtataccaaagtgaataaagtgtattccatccaaatltatgactccaagaacaaagac 1140
 QY 1141 aatgcatctgctgtatgaagcttgagaagcgtctgacttccaagacactgtgaacaa 1200
 Db 1141 aatgcatctgctgtatgaagcttgagaagcgtctgacttccaagacactgtgaacaa 1200
 QY 1201 gtgtgtctgcccacccagcgatga tgcgtgacagcagaacagctctgttgatttcggg 1260
 Db 1201 gtgtgtctgcccacccagcgatga tgcgtgacagcagaacagctctgttgatttcggg 1260
 QY 1261 tggggggccacccgagaggaagaagacactccaagaagtgtcgaacgctgcgaagtgctt 1320
 Db 1261 tggggggccacccgagaggaagaagacactccaagaagtgtcgaacgctgcgaagtgctt 1320
 QY 1321 cctcatggaacacagaagatgcaaacagcagatagttctatgaacaactgtatccaacagac 1380
 Db 1321 cctcatggaacacagaagatgcaaacagcagatagttctatgaacaactgtatccaacagac 1380
 QY 1381 atgatactgtgcccgtcttccctgcaagggaacgttcgaattccgcaagggtgaacgtgagg 1440
 Db 1381 atgatactgtgcccgtcttccctgcaagggaacgttcgaattccgcaagggtgaacgtgagg 1440
 QY 1441 cctctgtgtaacttgagaagaacataatctgtgtcgtatagagggaatacgaagctgggtctt 1500
 Db 1441 cctctgtgtaacttgagaagaacataatctgtgtcgtatagagggaatacgaagctgggtctt 1500
 QY 1501 ggcgtgtgccaaggtctatacgaacagagtgatacgggaatgtatgtatltcaggaactg 1560
 Db 1501 ggcgtgtgccaaggtctatacgaacagagtgatacgggaatgtatgtatltcaggaactg 1560
 QY 1561 attatctgacaaatgagggcagaagctaaatccacaatgcttccctgtgacgtcgtttt 1620
 Db 1561 attatctgacaaatgagggcagaagctaaatccacaatgcttccctgtgacgtcgtttt 1620
 QY 1621 tacaagaaaaaatgtgggtgtttgtgtctcccggtgacatgatttactcttgaagatgat 1680
 Db 1621 tacaagaaaaaatgtgggtgtttgtgtctcccggtgacatgatttactcttgaagatgat 1680
 QY 1681 tcagaggtcaactcatatttataaacaagtgaactgtctgacaaataaaaaaa 1738

|||||
DB 758 agcggatccaccagcttatgaactgaacaacagtcgagcatgtcatctataaa 817
OY 781 aaactgtaccacagtgatgctgttcttcaaaagcagtggttcttcaagctatagcc 840
DB 818 aaactgtaccacagtgatgctgttcttcaaaagcagtggttcttcaagctatagcc 877
OY 841 tggcgggtcaactgtgaactcaaacgcgcaagagcagatgttgccgagcagcagc 900
DB 878 tggcgggtcaactgtgaactcaaacgcgcaagagcagatgttgccgagcagcagc 937
OY 901 ccgagggtccttgccct-ggcaagtcagcctgcagctcccaagaacgctgcagcagc 959
DB 938 ccgagggtccttgcccttgagcagctgcagcctgcagcagcagcagcagcagcagc 997
OY 960 ctccatcatcaccctccagatgagatcgttgcaagccgcccactgcttggaataacctttaa 1019
DB 998 ctccatcatcaccctccagatgagatcgttgcaagccgcccactgcttggaataacctttaa 1057
OY 1020 caatcatcagcatctgagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1079
DB 1058 caatcatcagcatctgagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1117
OY 1080 agccggaataccaaagtagaanaagtgatcttctcaatcaatctgactccaagccaagaa 1139
DB 1118 agccggaataccaaagtagaanaagtgatcttctcaatcaatctgactccaagccaagaa 1177
OY 1140 caatgacatctgcgctgatatgaagctgcagaagcctctgacttcaagaacttaatgaacc 1199
DB 1178 caatgacatctgcgctgatatgaagcctctgacttcaagaacttaatgaacc 1237
OY 1200 aggtgtctctcccaaccagcagatgtctgcagcagcaagcagcctgcgtgagttccg 1259
DB 1238 aggtgtctctcccaaccagcagatgtctgcagcagcaagcagcctgcgtgagttccg 1297
OY 1260 gtggggggccacccgagagagagagagagagacctaagaagtgtcgaagcgtcccaagtgct 1319
DB 1298 gtggggggccacccgagagagagagagagagacctaagaagtgtcgaagcgtcccaagtgct 1357
OY 1320 tctcatttgagacacagatagcagcagcagatgtcttaagacaactgtctacacagc 1379
DB 1358 tctcatttgagacacagatagcagcagcagatgtcttaagacaactgtctacacagc 1417
OY 1380 catgactgtgcccgtctcctcagagggaagcgtcgaattctgcagagggtacagtgag 1439
DB 1418 catgactgtgcccgtctcctcagagggaagcgtcgaattctgcagagggtacagtgag 1477
OY 1440 gctctgtgactctcgaagaacaatatctgtgtgctgatataggggatacaagctggggttc 1499
DB 1478 gctctgtgactctcgaagaacaatatctgtgtgctgatataggggatacaagctggggttc 1537
OY 1500 tggctgtgcaaaagcttaccagacagagatgtacagggagatgtatgtatctacagcagc 1559
DB 1538 tggctgtgcaaaagcttaccagacagagatgtacagggagatgtatgtatctacagcagc 1597
OY 1560 gatttctgacaataagagcagacagcgtatccacatggtctcctctgagcagtc 1619
DB 1598 gatttctgacaataagagcagacagcgtatccacatggtctcctctgagcagtc 1657
OY 1620 ttacaagaanaaacaatggggcgtgttctgtctcccgctgcaatgtatctcttgaagatga 1679
DB 1658 ttacaagaanaaacaatggggcgtgttctgtctcccgctgcaatgtatctcttgaagatga 1717
OY 1680 ttcaagaggtcactcttcttataaacaagtgactgtgtgc 1723
DB 1718 ttcaagaggtcactcttcttataaacaagtgactgtgtgc 1761

RESULT 4
AAZ90478
ID AAZ90478 standard; cDNA; 2479 BP.
XX
AC AAZ90478;

XX 06-JUN-2000 (first entry)
DT
XX
DE Ovr115 homolog protein encoding cDNA.
XX
XX CSG; cancer specific gene; cancer; gynecologic cancer; ovarian; breast;
XX endometrial; uterine; lung; cytotoxic; ss.
XX
XX Homo sapiens.
XX
XX WO200012758-A1.
XX
XX 09-MAR-2000.
XX
XX 01-SEP-1999; 99WO-US19655.
XX
XX 02-SEP-1998; 98US-0098880.
XX
XX (DIAD-) DIADEXUS LLC.
XX
XX Salceda S, Sun Y, Recipon H, Caferkey R;
XX
XX WPI; 2000-256657/22.
XX
XX P-PSDB; AAY57280.
XX
PT Diagnosing, staging, monitoring, imaging and treating cancer especially
PT gynecological cancers e.g. breast, ovarian cancer and lung cancer,
PT involves measuring cancer specific gene levels in cells and body fluids
PS
PS Claim 9; Page 49-50; 58pp; English.
XX
XX The invention relates to detecting, diagnosing metastasis and staging
XX cancer by measuring levels of cancer specific genes (CSG) in cells,
XX tissues or body fluids. Their remission and progression, decreases and
XX increases in CSG levels, is also monitored, by periodic sample analysis.
XX The methods are useful for detecting cancers, especially gynecologic
XX cancers which include ovarian, breast, endometrial and uterine cancer
XX and lung cancer. Antibodies against the CSGs labeled with paramagnetic
XX ions or a radioisotope is useful for imaging cancer and when conjugated
XX with a cytotoxic agent are useful for treating cancer. The present
XX sequence represents a Ovr115 homolog protein encoding cDNA, that
XX can be used for the detection of the various cancers.
XX
XX Sequence 2479 BP; 578 A; 650 C; 677 G; 574 T; 0 other;
SQ
Query Match 95.2%; Score 1655.2; DB 21; Length 2479;
Best Local Similarity 99.5%; Pred. No. 0;
Matches 1660; Conservative 0; Mismatches 8; Indels 0; Gaps 0;
OY 56 gtacatattgaacattccagatcactcactactgactgactgttgaatacagcagaatgg 115
DB 1 gtacatattgaacattccagatcactcactactgactgactgttgaatacagcagaatgg 60
OY 116 ctgtgaactaagggttaacccacagctatctgacctactctgaataacacctgatatcaaac 175
DB 1 ctgtgaactaagggttaacccacagctatctgacctactctgaataacacctgatatcaaac 120
OY 176 cggaaaaccctatcccgacacagccactgtgtgtcccaactgtctacgaagtgcattccg 235
DB 121 cggaaaaccctatcccgacacagccactgtgtgtcccaactgtctacgaagtgcattccg 180
OY 236 ctgaagtactaccgctcccggtgcccagtagcagccgagaggtctcgaagcaggtctca 295
DB 181 ctgaagtactaccgctcccggtgcccagtagcagccgagaggtctcgaagcaggtctca 240
OY 296 acccggtgctcgaagcagcccaaatcccatcccgaggagagtgtagtaactcaaaagacta 355
DB 241 acccggtgctcgaagcagcccaaatcccatcccgaggagagtgtagtaactcaaaagacta 300
OY 356 agaagactgtgcatcacttgaccctggagccttccctcctgtggagctgctgtgcg 415
|||||

```

Db 301 agaaagcactgtgcatcacccttgaggacccttcctcgttgaggcgtgcgtgcgcg 360
Oy 416 cggccactctcggaaattcatcggcagaagaatgtctccaactctggatagatgagct 475
Db 361 cgggcttaactcggaaattcatcggcagaagaatgtctccaactctggatagatgagct 420
Oy 476 cctcaggtacatcgtcaaacccctcctaacttgatgtatgagcgtgtcaacatgcgcgcg 535
Db 421 cctcaggtacatcgtcaaacccctcctaacttgatgtatgagcgtgtcaacatgcgcgcg 480
Oy 536 gggaggaagcagaatcgggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 595
Db 481 gggaggaagcagaatcgggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 540
Oy 596 cctccagaagagaaatcctcgtgacccctgtgtgcacaagaacagacgtgaaacgaaatcagcgc 655
Db 541 cctccagaagagaaatcctcgtgacccctgtgtgcacaagaacagacgtgaaacgaaatcagcgc 600
Oy 656 gggcggcctgcagaggacatgggtctataagaataatttctaactctagccaagaaatagtg 715
Db 601 gggcggcctgcagaggacatgggtctataagaataatttctaactctagccaagaaatagtg 660
Oy 716 atgacaagcggatcccaacagcttatagaataatgaaacagatgcggcaatgtcgaatact 775
Db 661 atgacaagcggatcccaacagcttatagaataatgaaacagatgcggcaatgtcgaatact 720
Oy 776 ataaaaacatgacacacagatgatacctgtctctcaaaaacagatggtcttcttaacgtgta 835
Db 721 ataaaaacatgacacacagatgatacctgtctctcaaaaacagatggtcttcttaacgtgta 780
Oy 836 tagcctgcggggatcaaatgtaaatcgaagcgcgcagagacagatgtgtggcggcgagagcg 895
Db 781 tagcctgcggggatcaaatgtaaatcgaagcgcgcagagacagatgtgtggcggcgagagcg 840
Oy 896 cgtcccgaggggcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 955
Db 841 cgtcccgaggggcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 900
Oy 956 gaggctcccatcacaaccccgagtgatcgtgacagcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1015
Db 901 gaggctcccatcacaaccccgagtgatcgtgacagcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 960
Oy 1016 ttaacaatccatgacatgtgacatgtgacatgtgacatgtgacatgtgacatgtgacatgt 1075
Db 961 ttaacaatccatgacatgtgacatgtgacatgtgacatgtgacatgtgacatgtgacatgt 1020
Oy 1076 atggagccggatatacacaagaatgaaatgaaatgaaatgaaatgaaatgaaatgaaatgaa 1135
Db 1021 atggagccggatatacacaagaatgaaatgaaatgaaatgaaatgaaatgaaatgaaatgaa 1080
Oy 1136 agaaacaatgacatgtgacatgtgacatgtgacatgtgacatgtgacatgtgacatgtgac 1195
Db 1081 agaaacaatgacatgtgacatgtgacatgtgacatgtgacatgtgacatgtgacatgtgac 1140
Oy 1196 aaccagctgtgtctgcacacacacacacacacacacacacacacacacacacacacacacac 1255
Db 1141 aaccagctgtgtctgcacacacacacacacacacacacacacacacacacacacacacacac 1200
Oy 1256 cggggctgggggacacacacacacacacacacacacacacacacacacacacacacacacac 1315
Db 1201 cggggctgggggacacacacacacacacacacacacacacacacacacacacacacacacac 1260
Oy 1316 tgcctcctcatctgagacacagagatgcaaacagacagatagctcctcaaacacacacacac 1375
Db 1261 tgcctcctcatctgagacacagagatgcaaacagacagatagctcctcaaacacacacacac 1320
Oy 1376 cagccatgacatcgtgtgcggctcctcgcagggagacgtcgaatcctcgcaggggtgcagctg 1435
Db 1321 cagccatgacatcgtgtgcggctcctcgcagggagacgtcgaatcctcgcaggggtgcagctg 1380
Oy 1436 gaggggcctctgtgacatcctcgaagaaacaataatcgtgtgagtgatgaggggtataaagctgg 1495
Db 1381 gaggggcctctgtgacatcctcgaagaaacaataatcgtgtgagtgatgaggggtataaagctgg 1440

```

```

Oy 1496 gtctcgtcgtgtgcacaaagcttacaagacagagatgtacgggaatgtatgtatcacgg 1555
Db 1441 gtctcgtcgtgtgcacaaagcttacaagacagagatgtacgggaatgtatgtatcacgg 1500
Oy 1556 actggaattatcgaacaatgagggcagaagcgtcaatccaatgtcctcgtctcgtgaagt 1615
Db 1501 actggaattatcgaacaatgagggcagaagcgtcaatccaatgtcctcgtctcgtgaagt 1560
Oy 1616 cgtttacaagaagaacatggggcgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1675
Db 1561 cgtttacaagaagaacatggggcgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1620
Oy 1676 atgattcagaaggtcactcatttattataaagctgaactgtctggc 1723
Db 1621 atgattcagaaggtcactcatttattataaagctgaactgtctggc 1668

RESULT 5
AAZ87813
ID AAZ87813 standard; DNA; 2479 BP.
XX
AC AAZ87813;
XX
DT 12-MAY-2000 (first entry)
XX
DE Human tumour suppressor TPMSR2 gene (GenBank Accn No: U75329).
XX
KM Tumour suppressor gene; TPMSR2; cancer; human; drug design;
KM gene therapy; protein therapy; ds.
XX
OS Homo sapiens.
XX
PN WO20000605-A1.
XX
PD 06-JAN-2000.
XX
PF 29-JUN-1999; 99WO-US14622.
XX
PR 29-JUN-1998; 98US-0091044.
XX
PA (MYRI-) MYRIAD GENETICS INC.
XX
PI Wong AKC, Tavligian SV, Teng DHF;
XX
DR WPI: 2000-170914/15.
XX
PT Novel tumor suppressor TPMSR2 used for the diagnosis and prognosis of
PI human cancer -
XX
PS Example 1; Page 84-85; 89pp; English.
XX
CC The invention provides a new tumour suppressor gene, designated TPMSR2.
CC The TPMSR2 polynucleotides and polypeptides can be used in methods for
CC diagnosing and prognosing predisposition to cancer in humans. The
CC polypeptides may also be used in assays to screen for compounds with
CC anti-cancer or therapeutic properties. The polypeptides are also useful
CC for rational drug design. The TPMSR2 polynucleotides and polypeptides
CC may be used for gene therapy and protein therapy. The present sequence
CC represents the TPMSR2 gene (Genbank Accn No: U75329).
XX
SQ Sequence 2479 BP; 578 A; 650 C; 677 G; 574 T; 0 other;

Query Match 95.2%; Score 1655.2; DB 21; Length 2479;
Best Local Similarity 99.5%; Pred. No. 0;
Matches 1660; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

Oy 56 gtcatatgaaatccagatatacattactcgtatgctgtatataacagcaagatgg 115
Db 1 gtcatatgaaatccagatatacattactcgtatgctgtatataacagcaagatgg 60
Oy 116 ctltgaaactcagggtcaccacacagctatgtgaccttactactatgaaaaacatgataccaac 175

```

|||||
Db 61 cttgagactcagggtacacacacagctatgtgacctactactatgaataacatggtatcaaac 120
QY 176 cggaaaaacccctatcccgacagccacatgtgtgtcccccactgtctacagagtccacgg 235
Db 121 cggaaaaacccctatcccgacagccacatgtgtgtcccccactgtctacagagtccacgg 180
QY 236 ctgaagactacccgtcccggtgtcccgatgaccccgaggtgtctgacgcgaagcttcca 255
Db 181 ctgaagactacccgtcccggtgtcccgatgaccccgaggtgtctgacgcgaagcttcca 240
QY 296 acccgtctgtcagacgagcccaaatcccccacccgagacagtgctgacccccaagacta 355
Db 241 acccgtctgtcagacgagcccaaatcccccacccgagacagtgctgacccccaagacta 300
QY 356 agaaagcaactgtgcatcaacctgaacctctggggaacctctcgtgtggagctgtgcgcg 415
Db 301 agaaagcaactgtgcatcaacctgaacctctggggaacctctcgtgtggagctgtgcgcg 360
QY 416 ctgagctactctggaagtctcatgagcgaagtgtctccaactcttgagatagatgagct 475
Db 361 ctgagctactctggaagtctcatgagcgaagtgtctccaactcttgagatagatgagct 420
QY 476 cctcaggtactctgcatcaaacccctcctaactgtgtgatggcggtgtaacacgtgccggcg 535
Db 421 cctcaggtactctgcatcaaacccctcctaactgtgtgatggcggtgtaacacgtgccggcg 480
QY 536 gggaggaacgaagaatcggtgtgtgtgcctctacgacccaactctcctcaagtgtaact 555
Db 481 gggaggaacgaagaatcggtgtgtgtgcctctacgacccaactctcctcaagtgtaact 540
QY 596 catctcagaggaagtctctgacccctgtgtgccaagaacgtctgagcagagaatacggcg 655
Db 541 catctcagaggaagtctctgacccctgtgtgccaagaacgtctgagcagagaatacggcg 600
QY 656 gggcggtctgacgagacatggtgtgtatagaataatttctaacttaagccaagatagtg 715
Db 601 gggcggtctgacgagacatggtgtgtatagaataatttctaacttaagccaagatagtg 660
QY 716 atgacagcggatccacacatttataaactgaaacaaagtgtccggcaatgtctgatact 720
Db 661 atgacagcggatccacacatttataaactgaaacaaagtgtccggcaatgtctgatact 720
QY 776 ataaaaactgtatccacatgtatgctgttcttcaaaaagatggttcttcaacgtgta 835
Db 721 ataaaaactgtatccacatgtatgctgttcttcaaaaagatggttcttcaacgtgta 780
QY 836 tagctgtcggtgtcaactgtgaactcaagccgacagacgagatgtgtggcgcgagagcg 895
Db 781 tagctgtcggtgtcaactgtgaactcaagccgacagacgagatgtgtggcgcgagagcg 840
QY 896 cgctcccggtgggtcgctgtgcaagtcaagctcagtcacgtccacgtgtgcg 955
Db 841 cgctcccggtgggtcgctgtgcaagtcaagctcagtcacgtccacgtgtgcg 900
QY 956 gaggtccatcatcaccccgagttgatcggtgacacggccactcggtggaaaaacctc 1015
Db 901 gaggtccatcatcaccccgagttgatcggtgacacggccactcggtggaaaaacctc 960
QY 1016 ttaacaatcatgtgcatgtgacggtcatgtgcggggaatttgaagacatcttcatgtct 1075
Db 961 ttaacaatcatgtgcatgtgacggtcatgtgcggggaatttgaagacatcttcatgtct 1020
QY 1076 atggggtcggtatccacagatagaataagtgatcttcatccaatttgatccaaagca 1135
Db 1021 atggggtcggtatccacagatagaataagtgatcttcatccaatttgatccaaagca 1080
QY 1136 agaaacatatactgtgcgtatgaagtctgacgaagctctgacttcaagacactgtga 1195
Db 1081 agaaacatatactgtgcgtatgaagtctgacgaagctctgacttcaagacactgtga 1140
QY 1196 aacagatgtctgtcccaacccagatgatgtgtgacgacagacagctgtgtgattt 1255
|||||

Db 1141 aacagatgtctgtcccaacccagatgatgtgtgacgacgaagacagctgtgtgattt 1200
QY 1256 cgggtgtggggggccacccgagagaagaagagacctaagaagtgtctaaagctgtccaa 1315
Db 1201 cgggtgtggggggccacccgagagaagaagagacctaagaagtgtctaaagctgtccaa 1260
QY 1316 tgccttcatcttgagaacacagagatgtgacacagcagatgtgtctatgaacactgaacac 1375
Db 1261 tgccttcatcttgagaacacagagatgtgacacagcagatgtgtctatgaacactgaacac 1320
QY 1376 cagccatgatctgtgccggtctctcagagggaacgtcgaattcttgcagaggtgacagtg 1435
Db 1321 cagccatgatctgtgccggtctctcagagggaacgtcgaattcttgcagaggtgacagtg 1380
QY 1436 gagggcctctgttcaactctgaagaacaatatctgtgtgtgataaggagataacagctgg 1495
Db 1381 gagggcctctgttcaactctgaagaacaatatctgtgtgtgataaggagataacagctgg 1440
QY 1496 gtctgtgtgtgtcccaagactacagaccagagatgtgacgggaatgtgatatctacgg 1555
Db 1441 gtctgtgtgtgtcccaagactacagaccagagatgtgacgggaatgtgatatctacgg 1500
QY 1556 actggaattatcgcacaaatgaaggcagacgctaataccacatgtctcgtctcgtgaact 1615
Db 1501 actggaattatcgcacaaatgaaggcagacgctaataccacatgtctcgtctcgtgaact 1560
QY 1616 cgtttcaagaagaacaatggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1675
Db 1561 cgtttcaagaagaacaatggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1620
QY 1676 atgattcagaggtcacttacttatttaaacagtgtaactgtgtgtgt 1723
Db 1621 atgattcagaggtcacttacttatttaaacagtgtaactgtgtgtgt 1668
|||||

RESULT 6
AAC83325
ID AAC83325 standard; DNA; 3966 BP.
XX
AC AAC83325;
XX
DT 26-FEB-2001 (first entry)
XX
DE Human TMPRSS2 DNA.
XX
KW Prostate specific androgen regulated protein; ARSDRL; TMPRSS2;
KW PART-1; neoplastic; ds.
XX
OS Homo sapiens.
XX
PN MO200065067-A2.
XX
PD 02-NOV-2000.
XX
PF 21-APR-2000; 2000MO-US10920.
XX
PR 23-APR-1999; 99US-0130778.
PR 30-AUG-1999; 99US-0151585.
PR 30-DEC-1999; 99US-0174003.
PR 24-JAN-2000; 2000US-0177751.
XX
XX
XX (UNIM) UNIV WASHINGTON.
XX
PI Nelson PS, Hood L, Lin B;
XX
DR MPI: 2000-679676/56.
XX
DR P-PSDB; AAB36901.
XX
PT Polynucleotide encoding prostate specific androgen regulated
PT polypeptides and inhibitor of the peptides useful for treating or
PT reducing the progression of prostate neoplastic condition in an
PT individual -
XX

PS Claim 1; Page 81-86; 121pp; English.
XX The present invention relates to prostate specific androgen regulated
CC proteins. The invention may be used to determine an expression level
CC of the prostate-specific proteins ARSDR1, TMPRSS2, or PAPF-1 in a
CC fluid sample or prostate cell sample from an individual. It may also
CC be used for diagnosing and predicting the susceptibility of a
CC prostate neoplastic condition in an individual. Inhibitors of the
CC proteins are useful for treating or preventing the progression of a
CC prostate neoplastic condition.
XX
SQ Sequence 3966 BP; 992 A; 998 C; 988 G; 984 T; 4 other:

Query Match 95.2%; Score 1655.2; DB 21; Length 3966;
Best Local Similarity 99.5%; Pred. No. 0;
Matches 1660; Conservative 0; Mismatches 8; Indels 0; Gaps 0;
QY 56 gtcataatgaaatcctcagatcctatcattacatcagatgctgataaagaagatg 115
Db 1 gtcaatctgaaatcctcagatcctatcattacatcagatgctgataaagaagatg 60
QY 116 ctttgaatcagagtgcaacacagaggtatgagccttactatgaataaacatgatacaac 175
Db 61 ccttgaaactcagagtgcaacacagaggtatgagccttactatgaataaacatgatacaac 120
QY 176 cggaaaaccccatcctccgacacagcccaactggtccccaactgtctaaagagtgatcccg 235
Db 121 cggaaaaccccatcctccgacacagcccaactggtccccaactgtctaaagagtgatcccg 180
QY 236 ctcaatgaataccggtcccggtgcccagtaacgcccgaagggtccttgacgaggtctcca 295
Db 181 ctcaatgaataccggtcccggtgcccagtaacgcccgaagggtccttgacgaggtctcca 240
QY 296 accccgctgctgcacgacgagcccaatccccaatcccgagagtgagacctaagaagacta 355
Db 241 accccgctgctgcacgacgagcccaatccccaatcccgagagtgagacctaagaagacta 300
QY 356 agaaagcaactgtgcatcactccttgaaaccttgaggacctctcctgaggaggtgctgaccg 415
Db 301 agaaagcaactgtgcatcactccttgaaaccttgaggacctctcctgaggaggtgctgaccg 360
QY 416 ctgagcctactcctggaagtctcatgtggcagaagtgctccaactctgggataaggtctcgact 475
Db 361 ctgagcctactcctggaagtctcatgtggcagaagtgctccaactctgggataaggtctcgact 420
QY 476 cctcaggtacactgcatcacaacccctcctaactgtgtgcatgagctgtgcaacatgcccggcg 535
Db 421 cctcaggtacactgcatcacaacccctcctaactgtgtgcatgagctgtgcaacatgcccggcg 480
QY 536 gggaggaagaaatcagtggtgtctgcctctcaagacaaactctccttcagtgact 595
Db 481 gggaggaagaaatcagtggtgtgtctgcctctcaagacaaactctccttcagtgact 540
QY 596 catctcagaagaagtcctcggcaacctgtgtgccaagaagactctgaaagaaactcaggcg 655
Db 541 catctcagaagaagtcctcggcaacctgtgtgccaagaagactctgaaagaaactcaggcg 600
QY 656 gggcgagctgcaaggagacatgagcttaagaataatcttactcctgagcgaagaatagtg 715
Db 601 gggcgagctgcaaggagacatgagcttaagaataatcttactcctgagcgaagaatagtg 660
QY 716 atgacagcgaatccacagcttatagaatactgaaacaagtgccgcaatgtcgaatct 775
Db 661 atgacagcgaatccacagcttatagaatactgaaacaagtgccgcaatgtcgaatct 720
QY 776 ataaaaaactgtacacacagtgatgctgtctctcaaaagcagtggtcttcttaagctgtta 835
Db 721 ataaaaaactgtacacacagtgatgctgtctctctcaaaagcagtggtcttcttaagctgtt 780
QY 836 tagcctgaggggtcaacttgaactcaaaagccgacagagaggttctggcgcgacagagcg 895
Db 781 tagcctgaggggtcaacttgaactcaaaagccgacagagaggttctggcggttgagagcg 840

QY 896 cgtcccgaggggcctgccccttgacaggtcagcttgcacgttccagaacgtccacgtgtgcg 955
Db 841 cgtcccgaggggcctgccccttgacaggtcagcttgcacgttccagaacgtccacgtgtgcg 900
QY 956 gaggctcatcatcaccaccccgagtgatctcgatgaacgacgcccacactgcgttgaaaacctc 1015
Db 901 gaggctcatcatcaccaccccgagtgatctcgatgaacgacgcccacactgcgttgaaaacctc 960
QY 1016 ttaacaatccatgagatctggaacgcatctgcgggattcttgagacaatcttcaatgtctct 1075
Db 961 ttaacaatccatgagatctggaacgcatctgcgggattcttgagacaatcttcaatgtctct 1020
QY 1076 atggaacggatataccaagaatgaanaaagtattctcattccaataatatagtactccaagacca 1135
Db 1021 atggaacggatataccaagaatgaanaaagtattctcattccaataatatagtactccaagacca 1080
QY 1136 agaaacaatgacacttgctgtatgaagctgcagaaagcctctgaacttcaacgaactcagta 1195
Db 1081 agaaacaatgacacttgctgtatgaagctgcagaaagcctctgaacttcaacgaactcagta 1140
QY 1196 aaccagtgctgtccccaacccaggtcatgatatgtcagccagaacagctctgagattc 1255
Db 1141 aaccagtgctgtccccaacccaggtcatgatatgtcagccagaacagctctgagattc 1200
QY 1256 ccgggttgaggggccacccgagaggaagaaagacactcagaagtgtctgaacgttgcacaag 1315
Db 1201 ccgggttgaggggccacccgagaggaagaaagacactcagaagtgtctgaacgttgcacaag 1260
QY 1316 tgcttctcatatgagacacagaatgcaacacagagatagttcattagaacaactcgtatcacac 1375
Db 1261 tgcttctcatatgagacacagaatgcaacacagagatagttcattagaacaactcgtatcacac 1320
QY 1376 cagccatgactgtgcggtcctcctgcagggagaaagctcgaatctctgcagaagtgcacgtg 1435
Db 1321 cagccatgactgtgcggtcctcctgcagggagaaagctcgaatctctgcagaagtgcacgtg 1380
QY 1436 gagggcctctgtgtcactcctcgaagaacaatatctggtgagctgataaggatataagctggg 1495
Db 1381 gagggcctctgtgtcactcctcgaagaacaatatctggtgagctgataaggatataagctggg 1440
QY 1496 gttctgagctgtgccaagtcttaacagacagaggtgtaaggagatgtatgatatcagg 1555
Db 1441 gttctgagctgtgccaagtcttaacagacagaggtgtaaggagatgtatgatatcagg 1500
QY 1556 actgattatcatgacaatatgaagcgagacggttaataccaagtgcttcgtcctgaact 1615
Db 1501 actgattatcatgacaatatgaagcgagacggttaataccaagtgcttcgtcctgaact 1560
QY 1616 cgtttacaagaanaaacaatgaggctggtttgtcttcccgatgcatatctacttagag 1675
Db 1561 cgtttacaagaanaaacaatgaggctggtttgtcttcccgatgcatatctacttagag 1620
QY 1676 atgattcagaggttcaatcttatttaataaagaagaaactgtctggc 1723
Db 1621 atgattcagaggttcaatcttatttaataaagaagaaactgtctggc 1668

RESULT 7
AA287786
ID AA287786 standard; DNA; 1479 BP.
XX
AC AA287786;
XX
DT 12-MAY-2000 (first entry)
XX
DE Human tumour suppressor TMPRSS2 encoding DNA.
XX
XX Tumour suppressor gene; TMPRSS2; cancer; human; drug design;
KW gene therapy; protein therapy; ds.
XX
OS Homo sapiens.
XX

Key Location/Qualifiers
1..1479
/tag= a
/product= "TMPRSS2"
478
/tag= b
/note= "this base can be G or A with G being the more common allele. The codon will change from Val to Met"
625
/tag= c
/note= "this base can be T or A with T being the more common allele. The codon will change from Phe to Ile"
724
/tag= d
/note= "listed as T in GenBank Accn No: U75329"
768
/tag= e
/note= "this base can be C or T with C being the more common allele. This is a silent polymorphism"
777
/tag= f
/note= "this base can be C or T with C being the more common allele. The codon is unaffected with both alleles encoding Gly"
834
/tag= g
/note= "this base can be C or T with C being the more common allele. This is a silent polymorphism"
985
/tag= h
/note= "listed as C in GenBank Accn No: U75329"
1347
/tag= i
/note= "listed as C in GenBank Accn No: U75329"
1466
/tag= j
/note= "listed as A in GenBank Accn No: U75329"
1471
/tag= k
/note= "listed as A in GenBank Accn No: U75329"
MO200000605-A1.
06-JAN-2000.
29-JUN-1999; 99WO-US14622.
29-JUN-1998; 98US-0091044.
(MYRI-) MYRIAD GENETICS INC.
Wong AKC, Tavtigian SV, Teng DHF;
WPI: 2000-170914/15.
P-FSDB; AAY77726.
Novel tumor suppressor TMPRSS2 used for the diagnosis and prognosis of human cancer -
human cancer -
Claim 22; Page 77-79; 89pp; English.
The invention provides a new tumour suppressor gene, designated TMPRSS2. The TMPRSS2 polynucleotides and polypeptides can be used in methods for diagnosing and prognosing predisposition to cancer in humans. The polypeptides may also be used in assays to screen for compounds with anti-cancer or therapeutic properties. The polypeptides are also useful for rational drug design. The TMPRSS2 polynucleotides and polypeptides may be used for gene therapy and protein therapy. The present sequence represents the DNA encoding the TMPRSS2 polypeptide.

XX
SQ Sequence 1479 BP; 360 A; 406 C; 402 G; 306 T; 5 other:
Query Match 84.7%; Score 1472.4; DB 21; Length 1479;
Best Local Similarity 99.6%; Pred. No. 0;
Matches 1473; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
QY 112 atgctttagaactcaaggtgtacacacagctatttgacacctactatataaaacatgatac 171
DB 1 atgctttagaactcaaggtgtacacacagctatttgacacctactatataaaacatgatac 60
QY 172 caaccgaaaccccatcccgacagccacatggtgtccacactgttcaagagtgcat 221
DB 61 caaccgaaaccccatcccgacagccacatggtgtccacactgttcaagagtgcat 120
QY 232 ccgctcagttactaccctgtccctgtccacagtaacgcccggaggttctgaagcagct 291
DB 121 ccgctcagttactaccctgtccctgtccacagtaacgcccggaggttctgaagcagct 180
QY 292 tccaacccgctgtctgtcacagcagcccaatcccatccggagcagttgtcacctaaag 351
DB 181 tccaacccgctgtctgtcacagcagcccaatcccatccggagcagttgtcacctaaag 240
QY 352 actaagaagacactgtgcatcaccttgacaccttgaggacctctgttgaggctgctg 411
DB 241 nmmngaaagcactgtgcatcaccttgacaccttgaggacctctgttgaggctgctg 300
QY 412 gccctgtgacctctgtgaagttcatgtgagcagaatgtgtccaaacttgagatagtg 471
DB 301 gccctgtgacctctgtgaagttcatgtgagcagaatgtgtccaaacttgagatagtg 360
QY 472 gactcctcagtgactctgtacatcaaccccttaactgtgtgagtgtgtacacatgccc 531
DB 361 gactcctcagtgactctgtacatcaaccccttaactgtgtgagtgtgtgtacacatgccc 420
QY 532 ggcggggaagcagagatcggtgtgttcgctcttaaggacaaacttaactccttaagtg 591
DB 421 ggcggggaagcagagatcggtgtgttcgctcttaaggacaaacttaactccttaagtg 480
QY 592 tactcatctcagaggaagttctgtgacacctgtgtgccaagacagctggaagaaactac 651
DB 481 tactcatctcagaggaagttctgtgacacctgtgtgccaagacagctggaagaaactac 540
QY 652 ggcggggtgctgtcagcagagacatgggtgtatagaataatttttaactcttgccaagaa 711
DB 541 ggcggggtgctgtcagcagagacatgggtgtatagaataatttttaactcttgccaagaa 600
QY 712 gtggatgacagcagatccacacagcttatagaactgaacaaagtgccgcaatgtcgat 771
DB 601 gtggatgacagcagatccacacagcttatagaactgaacaaagtgccgcaatgtcgat 660
QY 772 atctataaaactgttaccacagtgatgctgttcttcaaaaagcagttgttcttaagc 831
DB 661 atctataaaactgttaccacagtgatgctgttcttcaaaaagcagttgttcttaagc 720
QY 832 tgtatagctgtgggtgtcaacttgaaactcaagcgcgcaagcagatgttggcggcgag 891
DB 721 tgtatagctgtgggtgtcaacttgaaactcaagcgcgcaagcagatgttggcggcgag 780
QY 892 agcgcgtcccgggggcctgtgacgttcagcttcagcttcagcttcagcttcagcttc 951
DB 781 agcgcgtcccgggggcctgtgacgttcagcttcagcttcagcttcagcttcagcttc 840
QY 952 tgcggagctccatcatcatcccccagtgatgtgtacagcgcgcccactgctgtgaaaaa 1011
DB 841 tgcggagctccatcatcatcccccagtgatgtgtacagcgcgcccactgctgtgaaaaa 900
QY 1012 cctcttaaaatccatgagatgtgagcagcattgtggggatgttggaataattttcatg 1071
DB 901 cctcttaaaatccatgagatgtgagcagcattgtggggatgttggaataattttcatg 960
QY 1072 ttctatgggcccgtatcatcaagaagatgatttctcatccaataatgatctccaag 1131

Db 682 ctaatlgagacagagatgcaacagcagatalgtctatgacaaacctgatacacacagcc 741
QY 1381 atgactcgttcggcgttcctccgacggggaacgttgatcttgccagggtgacggagg 1440
Db 742 atgactcgttcggcgttcctccgacggggaacgttgatcttgccagggtgacggagg 801
QY 1441 cctctgctcacttcgacgaacatattctgtgctgataaggaatacaagctgggtctc 1500
Db 802 cncctgctcacttcgacgaacatattctgtgctgataaggaatacaagctgggtctc 861
QY 1501 ggcctgtgcacaaagcttacagacagagagtgtaacgggaattgagtattcaaggacttg 1560
Db 862 ggcctgtgcacaaagcttacagacagagagtgtaacgggaattgagtattcaaggacttg 921
QY 1561 attatcgacaaatgagggagagcagcgttaacacatgcttcgtctcctgaagctgtt 1620
Db 922 attatcgacaaatgagggagagcagcgttaacacatgcttcgtctcctgaagctgtt 981
QY 1621 tacaagaaaacaatgagggagcgttttgctcccgctgacgatttactcttagagatgat 1680
Db 982 tacaagaaaacaatgagggagcgttttgctcccgctgacgatttactcttagagatgat 1041
QY 1681 tcagaggctcacttcattttatttaaacagtgaact 1715
Db 1042 tcagaggctcacttcattttatttaaacagtgaact 1076

RESULT 9

AAZ95004 standard; cDNA; 267 BP.

AAZ95004;

15-AUG-2000 (first entry)

CSG Prol15 EST useful as prostate cancer marker.

Prostate cancer; cancer specific gene; CSG; expressed sequence tag;

EST; diagnosis; monitoring; staging; imaging; therapy; metastasis;

marker; human; Prol15; ss.

Homo sapiens.

WO200023111-A1.

27-APR-2000.

19-OCT-1999; 99WO-US24331.

19-OCT-1998; 98US-0104737.

(DIAD-) DIADEXUS LLC.

Salceda S, Recipon H, Caferkey R;

WPI; 2000-339531/29.

Diagnosing, staging and monitoring the presence and metastases of

prostate cancer especially useful for treating prostate cancer

comprises measuring changes in cancer specific gene levels

Claim 7; Page 57; 74pp; English.

The present sequence is that of cancer specific gene (CSG) expressed
sequence tag clone 2189855H1 (Prol15); the full-length contig is
given in AAZ95005. The CSG was identified in a database search using
the data mining Cancer Leads Automatic Search Package (CLASP), which
allows the identification of highly expressed organ and cancer
specific genes. Overexpression of Prol15 was observed in 3 of 4
primary prostate cancer tissues examined, indicative of it being a
diagnostic marker for prostate cancer. The invention provides ESTs
and full-length contigs for prostate CSGs (see AAZ94998-295017). The
CSGs, polypeptides encoded by them, and antibodies that specifically

CC bind CSG are used in new, claimed methods for detecting, diagnosing,
CC monitoring, staging, imaging and treating prostate cancer. The new
CC methods provide earlier diagnosis for the presence and metastasis
CC of prostate cancer, and can be used to determine if a cancer has
CC metastasized, or to monitor the progress or stage of the disease
CC when it has not metastasized.

Sequence 267 BP; 64 A; 97 C; 53 G; 53 T; 0 other;

Query Match 15.4%; Score 267; DB 21; Length 267;
Best Local Similarity 100.0%; Pred. No. 66-66;
Matches 267; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 64 gaacattccagataccatattactctgactgctgtgataacagcaagatgcttgaac 123
Db 1 gaacattccagataccatattactctgactgctgtgataacagcaagatgcttgaac 60
QY 124 tcagggctacacacagcagctatggaaccttactatgaanaacatgataccaacggaaac 183
Db 61 tcagggctacacacagcagctatggaaccttactatgaanaacatgataccaacggaaac 120
QY 184 cccatcccgacagccacactgtgtccacactgtctctacagaggtgcatcccgctcagtlac 243
Db 121 cccatcccgacagccacactgtgtccacactgtctctacagaggtgcatcccgctcagtlac 180
QY 244 taccggtcccccgtgcccagtaagcccgaggggtcctgaacgagcttccaaacccgtc 303
Db 181 taccggtcccccgtgcccagtaagcccgaggggtcctgaacgagcttccaaacccgtc 240
QY 304 gctgcacgacgacccaatcccatcc 330
Db 241 gctgcacgacgacccaatcccatcc 267

RESULT 10

AAC06945

AAC06945 standard; cDNA; 272 BP.

AAC06945;

06-OCT-2000 (first entry)

Human secreted protein 5' EST, SEQ ID NO: 11020.

Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;

gene therapy; chromosome mapping; ss.

Homo sapiens.

EP1033401-A2.

06-SEP-2000.

21-FEB-2000; 2000EP-0200610.

26-FEB-1999; 99US-0122487.

(GEST) GENSET.

Dumas Milne Edwards J, Duclert A, Giordano J;

WPI; 2000-500381/45.

New nucleic acid that is a 5' expressed sequence tag (5' EST) for

obtaining cDNAs and genomic DNAs that correspond to 5' ESTs and for

diagnostic, forensic, gene therapy and chromosome mapping procedures

Claim 1; SEQ ID 11020; 71pp + CD-ROM; English.

The present sequence is one of a large number of 5' ESTs derived from

mRNAs encoding secreted proteins. No ORF has yet been conclusively

identified within the present sequence. The 5' ESTs were prepared from

Qy	811	aaagcaagtcgtcttccttcaagctctataagctcgcggggtcaactctgaactcaagccgcgaag	870
Db	714	ggccaacgtctggttaacactctgcagctgcacaagaccctctgtgtcatagaaagggtcttaacagctcac--	771
Qy	871	agcagagatctctggtgcagcagcagagacgcgtcccccgggggcctgcagccctctggcagaagtcacagcctc	930
Db	772	----gaaatcgtctgtgtgaaacaacatgctccctgtctctgcagatgagctccctgcagagccagcagctc	827
Qy	931	caagtcagagaagctcaacagctgtgtgtgtgtgtgtccatcatcaaccaccagatgtgactgtgtgaac	990
Db	828	caagcttcacagagctcaacacccctgtgcgggggcctctgtcatcaacacccctctgtgactcatcaact	887
Qy	991	ggccgcccaactctgtgtgaaaaaaacctcttaaacaaatccatgtgcactctgtgacagcgtactctgtgcggg	1050
Db	888	gtctgtcaacacgtgtctctgt	942
Qy	1051	attcttgagaaacaaatctctcaatgtctctatctggagccgggtatcaacagttagaagaaagtgtacttct	1110
Db	943	gtctagatctccctctgtgtgaaacatccaa--gtcccaatcccaactgtgtgtgagaaagtgtgtctaac	1001
Qy	1111	catccaaatctatgtactctccaaagaccaaagaaacatgtacatctgtgcctgtatgtaaagctgcagaag	1170
Db	1002	caacagcaagatcaacaaagacaaagagagcgtgtgtgcacatgtacatgtcccttatagactctgtgcgcgg	1061
Qy	1171	ccctctgtactcttcaacagcaccctagtgaaacaaagctgtgtctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt	1230
Db	1062	ccactcaccgtctcaatgaataatgtatcccaagcctctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt	1121
Qy	1231	caagcagaacaaagcaccctgtctgtgaattctccgggt	1290
Db	1122	cccgatcgtgtcaaaagagtgctgtgtgaagctcagatgt	1181
Qy	1291	tcgaagaatgtctgaaagcgcctgtgcagaagctgtctctctcatctgtgaacacaagaatgtgtcaacagcaga	1350
Db	1182	tcaccctgtccctgtgaacacaaagcgggcgcgtctccctctgtgaattctcaacaagaatctgtgacaacaaacgg	1241
Qy	1351	tatgtctcatatcaaacacactgtgaatcaacaaacaaacaaatgtatctgtgtgcgggcctctctcaaggggaac	1410
Db	1242	gagcgtgtgaagctgtgtgaacatcatctccctccctccatgtctctgtgcgggtgtctactctgtacgggtgtgc	1301
Qy	1411	gtctgaatctctgtccaaagctgtgaacagctgtgaagggcctctgtgtcactctcgaaagaaacataatctgtg	1470
Db	1302	gtgtgaacagcctgtccaaaggggtgaacagcgggggtccccctgt	1361
Qy	1471	tggt	1530
Db	1362	aaagt	1421
Qy	1531	tacggagagctgt	1590
Db	1422	tacaacccgtgtcaacccctctctctgt	1481
Qy	1591	tcacacatgt	1598
Db	1482	aactctgaag	1489

RESULT	13
AACT8475	
ID	AACT8475 standard; cDNA: 2412 BP.
XX	
AC	
XX	AACT8475;
DT	08-FEB-2001 (first entry)
XX	
DE	Human PRO382 (UNQ323) nucleotide sequence SEQ ID NO:68.
XX	
KW	Human: secreted protein; transmembrane protein; PRO; EST; cytosolic
KW	expressed sequence tag; detection; cancer; ss.
XX	
OS	Homo sapiens.
XX	

PN WO2000053756-A2.
 XX 14-SEP-2000.
 PD
 XX
 XX 18-FEB-2000; 2000OWO-US04341.
 XX
 PR 08-MAR-1999; 99WO-US05028.
 PR 12-MAR-1999; 99US-0123957.
 PR 29-MAR-1999; 99US-0126773.
 PR 21-APR-1999; 99US-0130232.
 PR 28-APR-1999; 99US-0131445.
 PR 14-MAY-1999; 99US-0134287.
 PR 23-JUN-1999; 99US-0141037.
 PR 26-JUL-1999; 99US-0145698.
 PR 29-OCT-1999; 99US-0162506.
 PR 30-NOV-1999; 99WO-US28313.
 PR 02-DEC-1999; 99WO-US28551.
 PR 02-DEC-1999; 99WO-US28565.
 PR 16-DEC-1999; 99WO-US30095.
 PR 30-DEC-1999; 99WO-US31243.
 PR 30-DEC-1999; 99WO-US31274.
 PR 05-JAN-2000; 2000OWO-US00219.
 PR 06-JAN-2000; 2000OWO-US00277.
 PR 06-JAN-2000; 2000OWO-US00376.
 XX
 PA (GETH) GENENTECH INC.
 XX
 PI Ashkenazi AJ, Baker KP, Botstein D, Desnoyers L, Eaton DL;
 PI Ferrara N, Filvaroff E, Fong S, Gao W, Gerber H, Gerritsen ME;
 PI Goddard P, Gudowski PJ, Grimaldi CJ, Gurney AL, Hillan KJ;
 PI Kijavitt JI, Kuo SS, Napier MA, Pan J, Paoni NF, Roy MA;
 PI Shelton DL, Stewart TA, Tumas D, Williams PM, Wood WI;
 XX
 DR WPI: 2000-611443/58.
 DR P-PSDB; AAB44250.
 XX
 XX Novel PRO polypeptides and polynucleotides used in detection methods,
 PT to target bioactive molecules to specific cells, and to modulate
 PT cellular activities -
 XX
 PS Claim 2; Fig 27; 636pp; English.
 XX
 CC AAC78458 to AAC78599 represent polynucleotide and EST (expressed
 CC sequence tag) sequences which encode secreted or transmembrane PRO
 CC polypeptides. The PRO polynucleotides and polypeptides have cytotostic
 CC activity. The polynucleotides and polypeptides can be used for detecting
 CC the presence of PRO polypeptides in samples, for linking bioactive
 CC molecules to cells and for modulating biological activities of cells,
 CC using the polypeptides for specific targeting. The polypeptide targeting
 CC can be used to kill the target cells, e.g. for the treatment of cancers.
 CC The polypeptide pairs provide specific targeting of bioactive molecules
 CC to cells. AAC78600 to AAC78987 represent PCR primers and probes used in
 CC the isolation of the PRO polynucleotide sequences.
 CC
 XX
 XX Sequence 2412 BP; 529 A; 648 C; 643 G; 592 T; 0 other;

Query Match	13.0%	Score 225.2;	DB 21;	Length 2412;
Best Local Similarity	54.4%;	Pred. NO. 1.2e-53;		
Matches 624;	Conservative	0;	Mismatches 488;	Indels 36;
			Gaps	7;
QY	469	tgcagctccctcagctacgtcatcaaccctctaactggtgtgacatgcgtgtcacactgc	528	
Db	360	tgtcgctatcctttaaigtatcgtcagctctgaagctgatatgtgaaggagctccgatatgc	419	
QY	539	cccgccggcggaagacagaaatcgtgtgttcgcctctacggaccacaacttccttcag	588	
Db	420	aaagcggcggaagacagctacccgtgtgtcgggtgtgtcagaatgcgtgtccag	479	
QY	559	gtgtactcatctcagagaagaatctccctgcgcacctgtgttcgaagaagacttggaaacgaac	648	
Db	480	gtgtcacaagctgcgtc-----tcgtggaaagacaaagtgtctccgaatggaagaggtctcc	533	

OY	649	tacggcgcgcgccgctgcgagggacgagcgtatagaagataatttactcttagccaaaga	708
Db	534	tacgcaaatgtctgcctgtgcccacactgtgggttctccaaagcttagtgcagataaacctc	593
OY	709	ataagctg_gatgacacgcgcgaatccacccagcttataagaaactgaaacaaagctgcgcgcaatgt	767
Db	594	agagctggagctctgcgtcgagggcgacgcttcgcggagagagttgtgtccatcatgacactcttg	653
OY	768	cgatatctataaaaaactgtaccac-----agtgtgcctgttcttca	810
Db	654	ccaagatgacaagagtgacatgcattaccacacactcagataatgtgagagagagatgtgcctct	713
OY	811	aaagcagatggtttctttacgcctgtatagcctctcggggtcgaacttgaactcaagcgccgacg	870
Db	714	ggccagctgtgttaactctgtcagctgcgaacgcctgtgttcataagaagggctgtaccagctcac--	771
OY	871	agcagagatgtgtggcgcgagagacgcgcgtccgcggggcctctgagcccttgccagctcagcctg	930
Db	772	----gcatcgtgggtgtgnaaaaactgatacctgtgctctgcagtgccctgcaagccagcctt	827
OY	931	caagctccaagaacgtccacgcgtgtcggaagctccatcatatacccccgaaatgtgatctgaca	990
Db	828	caattcccaagggctaccacccgcctgtcgggggcctcgtcatatacgcctcctgttgatactcct	887
OY	991	ggcccccacatgcgtgtgnaaaaactcttaacaatcagctatgacgtgacgcgcatcttgcgggg	1050
Db	888	gctgcacacatcgtgttatga--ctgttaactccccaagctcagtcagc---catccagctgg	942
OY	1051	attttgagacaactcttcattcgttctctatgtgagccggatatacgaatagaanaagttactct	1110
Db	943	gtctagattccctctgttgnaaatacca-gccccaatcccaacttgtgtgagaagaattgtctac	1001
OY	1111	catccaattatgtgctccacagacacaaagaacatgaaattgctgctgtatgaagctcagaag	1170
Db	1002	cacgcacagctacaagaagccaaagaaggtcgggcataatgacatgcgcctctatgaactgtgcggg	1061
OY	1171	cccttgaccttcaacgaacactagtgaacacagtgctgtctgcgcccacccagagatgagtctg	1230
Db	1062	ccactcactgcttcaatgaaatgatactcagcctgtgtgcctgcccacactctgaagagaatttc	1121
OY	1231	cagccagaaacgcgtctgtcgtgattctccgggtgtggggcgacccagagagagaagagaagacc	1280
Db	1122	cccgatgtgaaaagctgcgtcgtgacgttcaagatctgggtggggccacagaagatgtgaggtgtacgcc	1181
OY	1291	tcaagaagtgtctgaacgcgtctgccaaggtgtcttccatctgaaagacacagatgtcaacacgcga	1350
Db	1182	tcacctgtctcgaaacccaacgcggcgctgcctcttgatttccaanaagatctgtcaacacacgg	1241
OY	1351	tatgtctatgtacaacctgtatcacacaacgaatgatctgtgcgcgctcttcctgcaggaggaa	1410
Db	1242	gacgtgtacaggtgtgcatacatctccctccatcatctctgtcggcggtgactctgaacggtgtgc	1301
OY	1411	gtcattctcttggccaggggttgacagtggagggccctctgtcaactctgaaacaaatatatctgg	1470
Db	1302	gttgaaacgcgtcgccaggggggacagcgggggggcccctcgtgtgtgttcaagaagaagagctgtgg	1361
OY	1471	tgtcgtataaggggatacaaacgtcgggtgtctcgtcgtgtcacaagacttaagaacccaagagt	1530
Db	1362	aagtgtatgtgagcagacacacttgtgcacgtcgtcgtgcagaggtgaacaagcctgtgggt	1421
OY	1531	taggggaatttgtatgtattcacaagcgtctgattatcgacaatatgagcgacagcgctaa	1590
Db	1422	tacaccctgttcaacctctctcctgtgactgtgataccacgacgaaatgtgagagagactactaaa	1481
OY	1591	tccacatg 1598	
Db	1482	acctgaa 1489	
RESULT 14			
AAAG93842			
ID AAAG93842 standard; cDNA; 2413 BP.			
XX			

AC		AAA93842;
XX		
DT	11-JAN-2001	(first entry)
XX		Tumour associated differentially-expressed gene 12 cDNA sequence.
DE		
XX		
KM		Transmembrane serine protease; TADG-12; chromosome 17; vaccination; ss;
KW		tumour associated differentially-expressed gene 12; cytosolic; human;
XX		malignant hyperplasia; cancer; ovary; breast; lung; colon; prostate.
XX		
OS	Homo sapiens.	
PN	WO200052044-A1.	
XX		
PD	08-SEP-2000.	
XX		
PF	02-MAR-2000; 2000WO-US05612.	
XX		
PR	03-MAR-1999; 99US-0261416.	
XX		
PA	(UYAR-) UNIV ARKANSAS.	
PI	O'Brien TJ, Underwood LJ,	
DR	WPI: 2000-533263/48.	
XX	P-PADB: AAB32246.	
DR	DNA fragment encoding tumor associated differentially-expressed gene 12	
PT	protein used for diagnosing and treating malignant hyperplasia and	
PT	cancers including ovarian cancer -	
PS	Claim 2; Figure 4; 118pp; English.	
XX		
CC	This invention relates to a novel transmembrane serine protease called	
CC	tumor associated differentially-expressed gene 12 (TADG-12). TADG-12 is	
CC	located on chromosome 17. Sequences AAA93842-A93845 and AAB32246-B32249	
CC	represent human TADG-12 cDNA and their corresponding protein sequences.	
CC	A splice variant of TADG-12 (TADG-12V) leads to a truncated protein	
CC	product. TADG-12 is overexpressed in ovarian carcinomas. TADG-12	
CC	exhibits cytostatic activity, and can be used in vaccines and in gene	
CC	therapy. TADG-12 nucleotide and protein sequences are used in the	
CC	diagnosis of malignant hyperplasia and cancers or the ovary, breast,	
CC	lung, colon, prostate and other cancers where TADG-12 is overexpressed.	
CC	TADG-12 is particularly used as tumour marker for early disease	
CC	diagnoses. TADG12 proteins or fragments can be used to vaccinate an	
CC	individual with cancer, suspected of having a cancer or at risk of	
CC	getting cancer. Sequences AAA93846-A93853 represent PCR primers used for	
CC	amplifying the TADG-12 cDNA sequence, and in the quantitative analysis of	
CC	TADG-12 mRNA. AAB32250 represents a peptide fragment of TADG-12, used to	
CC	create anti-TADG-12 antibodies. Sequences AAB32251-B32369 represent	
CC	TADG-12 peptides which target HLA, and may be used in a vaccine or for	
XX	immune stimulation.	
SQ	Sequence 2413 BP; 554 A; 639 C; 636 G; 564 T; 0 other:	
Query Match	13.0%; Score 225.2; DB 21; Length 2413;	
Best Local Similarity	54.4%; Pred. No. 1.2e-53;	
Matches 624; Conservative	0; Mismatches 488; Indels 36; Gaps 7	
OY	469 tgcgactcctcaagtaacccaaccctaactgtgtgatggcgltcacatgc 528 DB 381 tgctgcatacccttaagtattgcgtcatataactcgtatgcaggatcttcgattgc 440 OY 529 ccgcgagggaagaagaagatcggtgtgttgcccctaaaggaccaaatcacttccctgg 588 DB 441 aaagacgagggaagacaagtagtacccgttgtgccgggtgggtgttcagaatgcgtgtccgcag 500 OY 589 gttctacttatccaaggaagatcctgcacacctgtgttcgaagacagctggaacgaac 648 DB 501 gggttcaaagcgtgc-----tcgtgggaagacatatgtgtccgaagactgggaaggtcac 554 OY 649 tacgggcgggcgacctgcaggaacatbggtcataaagaataattttactctaacgaaga 708	

```

Db      555 taagcaaatctgacctgctcccaactggcttcccaagctatgagttcagataaacctc 614
Oy      709 atagatg-gatgacagcgagtcacacagctttatgaaacgacagacagcgccgacatc 767
Db      615 agagtgagctcgcctcggaaggcggttcgggaagagttctgtctccatccatcaacctctg 674
Oy      768 cgaatctctaaaaaactctaccac-----agtgatctccgtctctca 810
Db      675 ccagatgacaaagtgacgtcattacacacactcagatatactgagggagggatgtgctctc 734
Oy      811 aaagcagtggtctctcttaccgtctatagcttcgggggtcaacttgaaacctcagccgacg 870
Db      735 ggcacagtggttacctctgacgtcagacagcctgtgtgtcatalagaaagggtcacagctcac-- 792
Oy      871 agcagagtggtggcgcgagagcgcgctcccggggaccttcggccctggcaggttcagctcg 930
Db      793 ----gcatcggtgggvggaacaatgctccttgctcctcgagtggtcctggagagccagccct 848
Oy      931 caagtcacagaagctccacgtgtgctgaggggtccatcatcaccccgagtgatctgtgaca 990
Db      849 cagttccaggggtcacacacgtgctggggggtctgtcatcacgcccctgtgatacact 908
Oy      991 gccgcccactgtgtggaataaacctcttaacatccatgagcatgtgacgagcatttcgggg 1050
Db      909 gctgcacacactgtgttatga--ctgtacctcccccagaatcaltgagc--catccaggtgg 963
Oy      1051 attctgagacaactctctctctatctatgagagcgagtcacaaagtgaataagtgattctc 1110
Db      964 gctctgcttccctgtctgagcaatcca-gcccatacccaacttggtggaagaagtgtctac 1022
Oy      1111 catcaaatatgactccacagaccaaagaacaatgacatgctgctgatagaaagctgcagaag 1170
Db      1023 caacagcaagtacaaagcaaaagagctgagcaatgacatgcaccttaagaaagctgcagg 1082
Oy      1171 cctctgaccttcaagagaccttaagaaacagtggtctgtgcccacacccagagcatgagctg 1230
Db      1083 ccaactacgttcaatgaatgaatgccagcctgtgtgcttcccaactctgaagaagaacttc 1142
Oy      1231 cagcagaacacagctcgtcgtggaattccgggtgagggcgacaggaagaagaaggaagacc 1290
Db      1143 ccggaatggaagaagtgctgctggagcgtcagaaatgaggggacacagaaatgaaagtacgctc 1202
Oy      1291 tcaagaagtgcgtgaacgctgcgaaggtgctctcatatgaaacacagaagatgcaacagcaga 1350
Db      1203 tcccctgtctgaaacacagcgcgctcccttctgatttccaaagaatctgtcaaacacag 1262
Oy      1351 tatgtctatgacaacctgattcacaccagcatgattctgtgcggctctctctgcaggggaac 1410
Db      1263 gacgtgtatgactgtgacatctctccctccatgctcgtcgcggtcactgacgtgggtg 1322
Oy      1411 gtcgaattctctcagaggtgacagtgagagggcctctgtcaactctgaaagaacaatatctg 1470
Db      1323 gtcgaacagctgcacagggagacagcggggcccctgtgtgtgtcagaagaaggaggtcgtgtg 1382
Oy      1471 tggctgataaggggatacaagctcgggtctctgtgctgtgcctcgaagagcttacagaccagagt 1530
Db      1383 aagctagtgaggcgagcagccttggtcagtcggcgtcgagaggtgaaacaagcctgggggtg 1442
Oy      1531 ttcgggaatgtagtattatccagatgattatctcgaatctgaaatgaaagcagcgtata 1590
Db      1443 taacaccgtgtacacctctctctgtgactgtgactccagacgaatggaagagaactctaa 1502
Oy      1591 tccacatg 1598
Db      1503 acctgaag 1510

```

```

XX      28-MAR-2001 (first entry)
DT      Human serine protease #4, encoding HCHAK72 cDNA clone.
XX
DE      Human
XX      Human, serine protease; osteopathic; immunosuppressive; antiallergic;
XX      antiinflammatory; cytosolic; cardiac; neuroprotective; nootropic;
XX      neuroleptic; vulnery; ophthalmological; antibacterial; antiviral;
XX      antifungal; antiparasitic; gene therapy; diagnosis; prevention; glaucoma;
XX      treatment; bone formation disorder; osteoporosis; arthritis; cancer;
XX      connective tissue disorder; autoimmune disorder; wound healing; asthma;
XX      systemic lupus erythematosus; male reproductive system disorder;
XX      testicular cancer; digestion and food absorption disorder; arhythmia;
XX      Crohn's disease; neurodegenerative disease; Alzheimer's disease; allergy;
XX      behavioral disorder; Tourette's syndrome; acute myelogenous leukemia;
XX      cardiovascular disorder; ocular disorder; drug screening; ss.
XX
OS      Homo sapiens.
XX
FH      Key      Location/Qualifiers
XX      CDS      201..1184
XX              /tag="a
XX              /product="Human serine protease #4 from clone HCHAK72"
XX
FT      WO200068247-A2.
XX
PD      16-NOV-2000.
XX
PF      05-MAY-2000; 2000WO-0512207.
XX
PR      07-MAY-1999; 99US-0133239.
XX      20-MAY-1999; 99US-0135163.
XX      03-AUG-1999; 99US-0147005.
XX      09-SEP-1999; 99US-0152935.
XX      01-NOV-1999; 99US-0162979.
XX
PA      (HUMA-) HUMAN GENOME SCT INC.
XX
PI      Ruben SM, Shi Y, Young PE, Ni J;
XX
DR      WPI: 2000-67979/66.
XX      P-PSDB: AAY72093.
XX
XX      New nucleic acid molecules encoding human serine protease polypeptides,
XX      useful for diagnosis, prevention and/or treatment of disorders e.g.
XX      osteoporosis, lupus erythematosus and Alzheimer's -
XX
PS      Claim 1: Page 272: 289pp; English.
XX
XX      The present cDNA sequence encodes human serine protease #4 from clone
XX      HCHAK72 (ATCC Deposit No: PTA27).
XX      It is useful in methods for the diagnosis, prevention and treatment of
XX      various disorders related to serine protease such as bone formation
XX      disorders (osteoporosis), connective tissue disorders (arthritis),
XX      autoimmune disorders (systemic lupus erythematosus), wound healing, male
XX      reproductive system disorders (testicular cancer), digestion and food
XX      absorption disorders (Crohn's disease), neurodegenerative diseases
XX      (Alzheimer's disease), behavioral disorders (Tourette's syndrome),
XX      proliferative and cancerous conditions (acute myelogenous leukemia),
XX      allergic reactions (asthma), cardiovascular disorders (arhythmia),
XX      ocular disorders (glaucoma) and infectious diseases caused by bacteria,
XX      viruses, fungi or parasites. It is also useful for screening therapeutic
XX      compounds. Serine proteases are used as immunological probes or
XX      tissues in biological samples, identification of male contraceptive
XX      agents, delivery of compositions to targetted cells expressing a
XX      receptor for serine protease, hybridisation probes and molecular weight
XX      markers. Serine protease nucleic acids are also useful in gene therapy.
XX
SQ      Sequence 1394 BP: 301 A; 381 C; 417 G; 295 T; 0 other:

```

• • • •

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 26, 2001, 09:21:33 : Search time 1348.83 Seconds
(without alignments)
12180.230 Million cell updates/sec

Title: US-09-615-285-1
Perfect score: 1738
Sequence: 1 ggcggagcgagcgagcgaggy.....ctggcaaaaaaaaaaaaaa 1738

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 10228115 seqs, 4726426750 residues

Total number of hits satisfying chosen parameters: 20456230

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*

1: gb_est1:*
2: gb_est2:*
3: gb_est3:*
4: gb_est4:*
5: gb_est5:*
6: gb_est6:*
7: gb_est7:*
8: gb_est8:*
9: gb_est9:*
10: gb_est10:*
11: gb_est11:*
12: gb_est12:*
13: gb_est13:*
14: gb_est14:*
15: gb_est15:*
16: gb_est16:*
17: gb_est17:*
18: gb_est18:*
19: gb_est19:*
20: gb_est20:*
21: gb_est21:*
22: gb_est22:*
23: gb_est23:*
24: gb_est24:*
25: gb_est25:*
26: gb_est26:*
27: gb_est27:*
28: gb_est28:*
29: gb_est29:*
30: gb_est30:*
31: gb_est31:*
32: gb_est32:*
33: gb_est33:*
34: gb_est34:*
35: gb_est35:*
36: gb_est36:*
37: gb_est37:*
38: gb_est38:*
39: gb_est39:*
40: gb_est40:*
41: gb_est41:*
42: gb_est42:*
43: gb_est43:*

44: em_esthum10:*
45: em_esthum11:*
46: em_esthum12:*
47: em_esthum13:*
48: em_esthum14:*
49: em_esthum15:*
50: em_esthum16:*
51: em_esthum17:*
52: em_esthum18:*
53: em_esthum19:*
54: em_esthum20:*
55: em_esthum21:*
56: em_esthum22:*
57: em_esthum23:*
58: em_esthum24:*
59: em_esthum25:*
60: em_esthum26:*
61: em_esthum27:*
62: em_esthum28:*
63: em_esthum29:*
64: em_esthum30:*
65: em_esthum31:*
66: em_esthum32:*
67: em_esthum33:*
68: em_esthum34:*
69: em_esthum35:*
70: em_esthum36:*
71: em_esthum37:*
72: em_esthum38:*
73: em_esthum39:*
74: em_esthum40:*
75: em_esthum41:*
76: em_esthum42:*
77: em_esthum43:*
78: em_esthum44:*
79: em_esthum45:*
80: em_esthum46:*
81: em_esthum47:*
82: em_esthum48:*
83: em_esthum49:*
84: em_esthum50:*
85: em_esthum51:*
86: em_esthum52:*
87: em_esthum53:*
88: em_esthum54:*
89: em_esthum55:*
90: em_esthum56:*
91: em_esthum57:*
92: em_esthum58:*
93: em_esthum59:*
94: em_esthum60:*
95: em_esthum61:*
96: em_esthum62:*
97: em_esthum63:*
98: em_esthum64:*
99: em_esthum65:*
100: em_esthum66:*
101: em_esthum67:*
102: em_esthum68:*
103: em_esthum69:*
104: em_esthum70:*
105: em_esthum71:*
106: em_esthum72:*
107: em_esthum73:*
108: em_esthum74:*
109: em_esthum75:*
110: em_esthum76:*
111: em_esthum77:*
112: em_esthum78:*
113: em_esthum79:*
114: em_esthum80:*
115: em_esthum81:*
116: em_esthum82:*

```
117: gb_est148:*
118: gb_est49:*
119: gb_est50:*
120: gb_est51:*
121: gb_est52:*
122: gb_est53:*
123: gb_est54:*
124: gb_est55:*
125: gb_est56:*
126: gb_est57:*
127: gb_est58:*
128: gb_est59:*
129: gb_est60:*
130: gb_est61:*
131: gb_est62:*
132: gb_est63:*
133: gb_est64:*
134: gb_est65:*
135: gb_est66:*
136: gb_est67:*
137: gb_est68:*
138: gb_est69:*
139: gb_est70:*
140: gb_est71:*
141: gb_est72:*
142: gb_est73:*
143: gb_est74:*
144: gb_est75:*
145: gb_est76:*
146: gb_est77:*
147: gb_est78:*
148: gb_est79:*
149: gb_est80:*
150: gb_est81:*
151: gb_est82:*
152: gb_est83:*
153: gb_est84:*
154: gb_est85:*
155: gb_est86:*
156: gb_est87:*
157: gb_est88:*
158: gb_est89:*
159: gb_est90:*
160: gb_est91:*
161: gb_est92:*
162: gb_est93:*
163: gb_est94:*
164: gb_est95:*
165: gb_est96:*
166: gb_est97:*
167: gb_est98:*
168: gb_est99:*
169: gb_est100:*
170: gb_est101:*
171: gb_est102:*
172: gb_est103:*
173: gb_est104:*
174: gb_est105:*
175: gb_est106:*
176: gb_est107:*
177: gb_est108:*
178: gb_est109:*
179: gb_est110:*
180: gb_est111:*
181: gb_est112:*
182: gb_est113:*
183: gb_est114:*
184: gb_est115:*
185: gb_est116:*
186: gb_est117:*
187: gb_est118:*
188: gb_est119:*
189: gb_est120:*
190: gb_est121:*
191: gb_est122:*
192: gb_est123:*
193: gb_est124:*
194: gb_est125:*
195: gb_est126:*
196: gb_est127:*
197: gb_est128:*
198: gb_est129:*
199: gb_est130:*
200: gb_est131:*
201: gb_est132:*
202: gb_est133:*
203: gb_est134:*
204: gb_est135:*
205: gb_est136:*
206: gb_est137:*
207: gb_est138:*
208: gb_est139:*
209: gb_est140:*
210: gb_est141:*
211: gb_est142:*
212: gb_est143:*
213: gb_est144:*
214: gb_est145:*
215: gb_est146:*
216: gb_est147:*
217: gb_est148:*
218: gb_est149:*
219: gb_est150:*
220: gb_est151:*
221: gb_est152:*
222: gb_est153:*
223: gb_est154:*
224: gb_est155:*
225: gb_est156:*
226: gb_est157:*
227: gb_est158:*
228: gb_est159:*
229: gb_est160:*
230: gb_est161:*
231: gb_est162:*
232: gb_est163:*
233: gb_est164:*
234: gb_est165:*
235: gb_est166:*
236: gb_est167:*
237: gb_est168:*
238: gb_est169:*
239: gb_est170:*
240: gb_est171:*
241: gb_est172:*
242: gb_est173:*
243: gb_est174:*
244: gb_est175:*
245: gb_est176:*
246: gb_est177:*
247: gb_est178:*
248: gb_est179:*
249: gb_est180:*
250: gb_est181:*
251: gb_est182:*
252: gb_est183:*
253: gb_est184:*
254: gb_est185:*
255: gb_est186:*
256: gb_est187:*
257: gb_est188:*
258: gb_est189:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

Db 49 TTATGACAATGAGGAGGAGGCTAATCATCAGATGCTTCCTTG 1

RESULT 2
 BB675496 814 bp mRNA EST 21-DEC-2000
 LOCUS 602138474F1 NIH_MGC_83 Homo sapiens cDNA clone IMAGE:4274658 5',
 DEFINITION mRNA sequence.
 BB675496
 ACCESSION BB675496.1 GI:11949391
 VERSION
 KEYWORDS
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 1 (bases 1 to 814)
 NIH-MGC http://mgc.nci.nih.gov/
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 TITLE
 JOURNAL
 COMMENT
 Email: cga@bs-remail.nih.gov
 Tissue Procurement: CLONETECH Laboratories, Inc.
 cDNA Library Preparation: CLONETECH Laboratories, Inc.
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 Plate: LCM1092 row: d column: 19
 High quality sequence stop: 568.
 Location/Qualifiers
 1. 814
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:4274658"
 /clone_lib="NIH_MGC_83"
 /lab_host="DH10B (T1 phage-resistant)"
 /note="Organ: prostate; Vector: pDMR-LIB (Clontech);
 Site_1: Sfil (98ccctcgccgc); Site_2: Sfil (98ccatctgccc
); 5' and 3' adaptors were used in cloning as follows: 5'
 adaptor sequence: 5'-CAAGGCGCATTTAGGCC-3' and 3' adaptor
 sequence: 5'-ATTCTAGAGCGCCGAGGCGCCGACATG-dT(30)BN-3'
 (where B = A, C, or G and N = A, C, G, or T). Average
 insert size 1.4 kb (range 0.5-4.0 kb). 14/15 colonies
 contained inserts by PCR. This library was enriched for
 full-length clones and was constructed by Clontech
 Laboratories (Palo Alto, CA)."
 BASE COUNT 210 a 177 c 257 g 170 t
 ORIGIN

Query Match 27.6%; Score 479.4; DB 168; Length 814;
 Best Local Similarity 90.3%; Pred. No. 1.4e-127;
 Matches 624; Conservative 0; Mismatches 56; Indels 11; Gaps 10;

638 ggaacgagaactacggcgccgctcgaggaactgggctataagaacttcttact 697
 Db 1 GGAACGAGAACTACGGCGCGCGCGC--TGCAGGACATGGGCTATAAATAATTTTACT 58
 638 ctaggcaagaaatagtagtagcagcaggtaccacagccttagaaactgaacacagag 757
 Db 59 CTAGCGAAGAAATAGTAGTAGACAGCGGATCCACCGCTTATGAACAGAACACAGATG 118
 758 ccggcaatgcatatctctaaacaaacttacacacagtgatgctctctcaaaagcag 817
 Db 119 CCGGCAATGTCATATCTATAAAACTGTACACAGTATGCTGTTTCAAAAGCAG 178
 818 tgggtctcttaccgtctatagcctcgagggtcaacttgaaactcaagcgcgacagcaga 877
 Db 179 TGGTTCTTTAGCGTGTATAGCCTCGGGGTCACACTTGAACCAAGCCGACAGACAGA 238
 878 ttgttgaggcgagagcgcgctccgggggcttgccctcgaggtcagctcgcacgctcc 937
 |||

Db 239 TTGTGGGCGCGAGAGCGCGCTCCGGGGGCGCTGCGAGGTGACCTGCACGTCC 298

938 agaagctccacgtgtgcgagagctccatcatcaccocccagtgagatctgacacgcc 997
 Db 299 AGAAGCTCCACCTGTGCGAGGCT-CATCATATACCCCGAGT-GATCGTAGACGCCGCC 356
 998 actgctgtgaaaaaactcttaacaaatccatgcatgcatgcatgcatgcatgcatgcat 1057
 Db 357 ACTGCGGTGAAAAACCTCTTACATATCATGATGATGATGATGATGATGATGATGATGAT 415
 1058 gacaaatcttctatgtctatgtagcagcagatcaccagatgaa-aagtgatcttcatca 1116
 Db 416 GACAACTTTTCATGTCTATATGAGGCGGATACCAAGTAAACAAAGTATTTCTCATCA 475
 1117 aattctgactcc-aagaccaaaacaaatcacat-tgcgctg-atgaactgcaagact 1173
 Db 476 AATTATGACTCCAAAGACCAACAAATACATGTCGCTGATGATGATGATGATGATGATGAT 535
 1174 ctgactctcaacgaact-agtgaaccagtgctct-gcccaaccagcagcatgactgctgc 1231
 Db 536 CTGACTTTCAACGACCTAATGTAACAGTGTGCTGGCCAAACAGGATGATGATGATGATG 595
 1232 agccgaagaagctctgctgacttccgggtgggggccaacgagagaagaaggaagact 1291
 Db 596 AGGCAAGAACAACTTCTGTGATTTCCGGGTGGGGCCACCGAGAGAGAGAAAGACTCC 655
 1292 cagaagtgtcgaagcgtgcgaagtgctct 1322
 Db 656 AGAAGTGTGTAAGTGTGAAGGGTTTAT 686

RESULT 3
 BE857738/3
 LOCUS 7647f08.x1 NCI CGAP_P128 Homo sapiens cDNA clone IMAGE:330639 3',
 DEFINITION similar to SW:TM82_HUMAN O15393 TRANSMEMBRANE PROTEASE, SERINE 2 ;,
 mRNA sequence.
 BE857738
 ACCESSION BE857738.1 GI:10372064
 VERSION
 KEYWORDS
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 1 (bases 1 to 465)
 NCI-CGAP http://www.nci.nih.gov/ncicgap.
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 Unpublished (1997)
 JOURNAL
 COMMENT
 Contact: Robert Strausberg, Ph.D.
 Email: cga@bs-remail.nih.gov
 Tissue Procurement: Michael J. Brownstein, M.D., Ph.D., Michael R.
 Emmert-Buck, M.D., Ph.D.
 cDNA Library Preparation: M. Bento Soares, Ph.D.
 DNA Sequencing by: Greg Lennon, Ph.D.
 Clone distribution: NCI-CGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL, send email to:
 info@image.llnl.gov
 Seg primer: -40bp from Gluco
 High quality sequence stop: 353.
 Location/Qualifiers
 1. 465
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:330639"
 /clone_lib="NCI_CGAP_P128"
 /sex="male"
 /dev_stage="adult"
 /lab_host="DH10B"
 /note="Organ: prostate; Vector: pT73D-Pac (Pharmacia)
 with a modified polylinker. Plasmid DNA from the
 normalized library NCI_CGAP_P128 was prepared, and ss

Db	121	AACGTCGATTTCTTCCAGAGGGGACAGTCGAGAGGGCCCTTGCTGCTTCGAGAACCAATTATC	180
Oy	1468	TCGTGTGCTGCATAGAGGGAGATACAAGCTGtggtctgcgtctgcacaaacttacaacacaga	152
Db	181	TGGTGGCTGCATAGGGGAGATACAGCTGTggttcttgctgtccCAAAGGTTTACACACAGGA	240
Oy	1528	gtctacgggaatgltgatggtatctcaagcactggaattatcgacaatatgagggacagagc	158
Db	241	GTGTACGGGAATGTGAATGGTATTCACCGAGCTGGATTATTCGACAAATTCAGAGCGACGCGC	300
Oy	1588	taatccacatggtcttcgtcccttcagctcgttttacaagaagaacaatgggcctgagtcttg	164
Db	301	TAAATCCAAATGGTCTTGCTGCTTGACCTGCTTTTACAGAAAACAAATGGGCGCTGGGCTTG	360
Oy	1648	cttcctccgtgcatagtattactctctagagatgattcagaggtcaacttcattt	1698
Db	361	CTTCCCGCTGCATATTACTCTTTAGAGATTCACAGAGCTCACTTCATTT	411

RESULT	7				
LOCUS	BF076876				
DEFINITION	BF076876	528 bp	mRNA	EST	18-OCT-2000
ACCESSION	226635	MARC	2BOV	Bos taurus	CDNA 5', mRNA sequence.
VERSION	BF076876				
KEYWORDS	BF076876.1	GI:10870706			
SOURCE	EST.				
ORGANISM	COW.				
	Bos taurus				

REFERENCE
AUTHORS
1 (bases 1 to 528)
Smith, T. P. L., Casas, E., Stone, R. T., Heaton, M. P., Grosse, W. M.,

TITLE Design and use of four pooled tissue normalized cDNA libraries for EST discovery in cattle
JOURNAL Unpublished (2000)
COMMENT Contact: Smith TPL
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA

Email: smulthe@mail.marc.usda.gov
Single pass sequencing. Bases called and alt trimmed with phred
v0.98b904.e. Vector identified by cross_match with the -mismatches
and -minmatch 12 options.
PCR Primers
FORWARD: AGGAACAGCTATGACCAT
BACKWARD: GTTTTCCAGTCACGACG
Plate: 88 row: N column: 4
Seq primer: ATTATGGTGACACCTATAG.

FEATURES	
source	location/qualifiers
	1. .528
	/organism="Bos taurus"
	/db_xref="taxon:9913"
	/clone_lib="MARC.2B0Y"
	/tissue_type="pooled"
	/lab_host="DH10B"
	/note="Vector: pCMV SPORT6; Site_1: XbaI; Site_2: XhoI;
	library made from pooled tissue from testis, thymus,
	semitendinosus muscle, longissimus muscle, pancreas,
	adrenal, and endometrium."
BASE COUNT	138 a 133 c 140 g 117 t
ORIGIN	

Query Match	23.0%;	Score 400.6;	DB 144;	Length 528;
Best Local Similarity	85.0%;	Pred. No. 8.1e-105;		
Matches 448; Conservative	0;	Mismatches 79;	Indels 0;	Gaps 0;
Oy	767	tcgatatctataaaaactgtaccacacgttgatgccctgtcttcttcaaaaagcagctggtttctt	826	

Db	2	TGGATCTCTATATAAAACCTACACAGGAGTGTGTTGTTCTTCAAAACACAGTGTTTCTT	61
Qy	827	taagctgtatagccttgcgggtgtaaccttgaactcaagccgcagagcaagattgttggcgc	886
Db	62	TACCGTGTATGAATAGCGGGGCTCTGGTGTAAGACAGCCGCCAGAGCCGGATGTGGGCG	121
Qy	887	gggaagagcgcgctcccgggggcgcttggcccttgcgaagtcaagcttgcagcttgcgaagcgttc	946
Db	122	GATCGAATGCTTACCTACGTGGCGGAATGGCTTGGCAGAGTTAGCTTGACAGTGACGGGCGATTC	181
Qy	947	acgtgtgcgaagctcccatcatcacccccgaatggatcgtgaacgcgcgcactgcgtgcg	1006
Db	182	ACGCTGTGGGGGCTCCATCATCACCCCGAGTGGATCTGACACCCCGCAGCTGCGTAG	241
Qy	1007	aaaaaccttacaatccatggcatctggaacgcatcttgcggggaatttggatacaatctt	1066
Db	242	AGGAACTCTTAAACATCCCAAGATCTGGGTGGCTTTCAGAGATCTTGAAACAACTTT	301
Qy	1067	tcatgtctataggaaccggaataccaagatgaaaaaagtgtattctcatccaatttgact	1126
Db	302	ACATTTCTTATGGAAGTGGATGCCAGTAGACAAAAGTGATTTTCCACCCAAATTTATGATT	361
Qy	1127	ccaagaccaagacatatgacattgcgcgtatgaagtgcagaagcctctgaacttcaagc	1186
Db	362	CCAAACCAAGACACATGACTTGGGCTATATAAAGTGCAGACGCCCTCTGACTTTCAACG	421
Qy	1187	acctagtgaaaacagttgttctgcccaaccgcagcatgatgtctgcagcaagaacagctt	1246
Db	422	ACAAAGTGAAGCCATGTGTCTTACCCCAATCCAGGATGATGCTGGAGCAACTCAGTCTT	481
Qy	1247	gctggaattccgggtggyggggccaccgggagaaaggaagaagcccca	1293
Db	482	GCTGGATTTTCCGGGTGGGGGGCCACTCTAGAAAGGGAAGACCTCCA	528

RESULT	8
LOCUS	A1673506/c
DEFINITION	A1673506 417 bp mRNA EST 18-DEC-1999 img9h04.xl Soares_Dlekgreife.colon.NHUC Homo sapiens cDNA clone IMAGE:2351093 3' similar to SW:TM52_HUMAN O13593 TRANSMEMBRANE PROTEASE, SERINE 2 ; , mRNA sequence.
ACCESSION	A1673506
VERSION	A1673506.1 GI:4853237
KEYWORDS	EST.
SOURCE	human.
ORGANISM	Homo sapiens

REFERENCE	(bases 1 to 417)
AUTHORS	NCI-CCAP http://www.ncbi.nlm.nih.gov/ncicgap .
TITLE	National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
JOURNAL	Tumor Gene Index
COMMENT	Unpublished (1997)
	Contact: Robert Strausberg, Ph.D.
	Email: cgapbs-r@mail.nih.gov
	This clone is available royalty-free through LINL; contact the
	IMAGE Consortium (infoimage.linl.gov) for further information.
	Insert Length: 1679 Std Error: 0.00
	Seq primer: -40UP from Gibco
FEATURES	High quality sequence stop: 215.
source	Location/Qualifiers
	1..417

```

/clone_lib="Soares.Dieckgrafe.colon.NHUC"
/tissue_type="colonic mucosa from 5 ulcerative colitis
patients"
/lab_host="DH10B (phage-resistant)"
/notes="Organ: colon; Vector: pT7TD-Pac (Pharmacia) with a
modified polylinker; Site:1: Not I; Site:2: Eco RI; 1st
strand cDNA was primed with a Not I - Site(cdr) primer [5'

```



```

Query Match          21.3%  Score 370.6; DB 19; Length 385;
Best Local Similarity 97.7%  Pred. No. 3.6e-96;
Matches 376; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

Oy 1201 gfgtctgcgccaaccagcagatgctgcgcagccagaacagctctgctggttccgg 1260
Db 385 GGTGTCTGTGCCCCACCCAGCATTAATGTGACCCAGAACACCTCTGTAATTTCCGGG 326
Oy 1261 tggggggccacagagagaagaagaccctcagaagtgctgaacgctgcgaagtgctt 1320
Db 325 TGGGGGGCCACCGAGAGAAAGGAAAGACTTAAAAAGTCTTAACCGCTCCCAAGGTGCTT 266
Oy 1321 ctcattgagacacagagatgcaacagcagatatgtctatgacaacctgatacaccaagcc 1380
Db 265 CTCATTAAAGACACAGAGATGCAACAGCATATGTCTATGACACACTGATCAGACAGCC 206
Oy 1381 atgactctgcgcggtctctctgcaggggaaagctcgaattcttgcacaggggtgacagtgaggg 1440
Db 205 ATGATTTTGTGCGCGCTTCTCTGACGGGGAACGTGATTTCTTGCCAGGCTGACAGTGGAGGG 146
Oy 1441 cctctgctactcgaagacacatactctgtgctgatatagggatatacaagctcgaggtctc 1500
Db 145 CCTCTGTGCTACTTCCAGAGACAATTTTGGTGTGCTGATAGGGGATACAACTGGGGTTCT 86
Oy 1501 ggcgtgcacaaagcttacagaccagagtgtaagggaaatgtaagtgatctacaggaactgg 1560
Db 85 GGCTGTGCCAAAGCTTACAGACAGAGAGTGTACGGGAATGTATGATTTACAGGACTGG 26
Oy 1561 attatgacaataatgaggcgagcg 1585
Db 25 ATTATCGACAAATGAGGCGACAGC 1

RESULT 12
LOCUS      BF857022      497 bp      mRNA      EST      16-JAN-2001
DEFINITION QV3-FT0210-031100-436-402 FT0210 Homo sapiens cDNA, mRNA sequence.
ACCESSION  BF857022
VERSION     BF857022.1 GI:12244766
KEYWORDS   EST.
SOURCE      human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE   1 (bases 1 to 497)
AUTHORS    Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Britones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., deoliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
            'M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
            Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
            Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
COMMENT     Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?cl=QV3&ct2=QV3-FT0210-
            031100-436-402&f3=2000-11-03&f4=1)
            Seq primer: puc 18 forward
            High quality sequence start: 7
            High quality sequence stop: 497.
            Location/Qualifiers
            1..497
FEATURES
source

```

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_1lb="FT0210"
/seq_stage="Adult"
/notes="Organ: prostate,tumor; Vector: puc18; Site_1: SmaI;
Site_2: SmaI; A mini-library was made by cloning products
derived from ORESTES PCR (U.S. Letters Patent application
No. 196,716 - Ludwig Institute for Cancer Research)
profiles into the puc 18 vector. Reverse transcription of
tissue mRNA and cDNA amplification were performed under
low stringency conditions."

BASE COUNT      104 a      130 c      137 g      126 t
ORIGIN

Query Match          21.1%  Score 366; DB 170; Length 497;
Best Local Similarity 98.4%  Pred. No. 8.5e-95;
Matches 380; Conservative 0; Mismatches 5; Indels 1; Gaps 1;

Oy 521 cacactgccccgcgggagagagaaatcggtgtgttcgcctcctacggaaccaactca 580
Db 403 CCCCGTCCCCCGGGGAGAGAGAAATCGGTGTTCGCTTACGAGCCAACTTCA 344
Oy 581 tccctcagtgtaactcattctcagaagaagtcctcgacccctgtgtgccaagacgactgga 640
Db 343 TCCCTTCAGATGTGATCACTCAGAGAAAGTCTTGACACCTGTGTGCCAAGACGACTGGA 284
Oy 641 acgagaactcagggcgggcgctcgaaggacatggtgcataagaataatttactcta 700
Db 283 ACGAGAACTACGGCGGGCGGCGCTCGACAGGACATGGGCTATAAGATATTTTACTCTA 224
Oy 701 gccaaagaaatgtagatgacagcgagatccacagcgttataagaaactgaacacaagtgcg 760
Db 223 GCCAAGAAATAGTGAATGACACAGCGATCCACGAGCTTATGAACTGAACACAAAGTGGCG 164
Oy 761 gcaatgcatatctataaaacctgtaccacagatgagatgagcgttctcaaaagaggg 820
Db 163 GCAATGTCATATCTATAAAACGTACACAGTGAATGCCGTTCTTCAAAAGCAAGTGG 104
Oy 821 ttctctacgctgtagctgtcggtggtcaacttgaactcaagccgacagacaga-tt 879
Db 103 TTTCTTACGCGTATPAGCGTGGGGGTCAACTTGAACCAAGCCGACAGACAGATT 44
Oy 880 gtggcgcgagagagcgcgctccggg 905
Db 43 GTGGCGGTGAGAGCGCGCTCCCGG 18

RESULT 13
LOCUS      BE483122      490 bp      mRNA      EST      28-AUG-2000
DEFINITION BE483122 BARC 5BOV Bos taurus cDNA 5', mRNA sequence.
ACCESSION  BE483122
VERSION     BE483122.1 GI:9602655
KEYWORDS   EST.
SOURCE      cow.
ORGANISM   Bos taurus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
            Bovidae; Bovinae; Bos.
REFERENCE   1 (bases 1 to 490)
AUTHORS    Sonstegard,T.S., Capuco,A.V., Van Tassel,C.P., Ashwell,M.S. and
            Wells,K.D.
            Mapping of Expressed Sequence Tags from a normalized bovine mammary
            gland cDNA library
            Unpublished (2000)
COMMENT     CONTACT: Sonstegard TS
            USDA, ARS, Beltsville Agricultural Research Center
            Bldg. 200 Rm 2A, Beltsville, MD 20705, USA
            Tel: 301 504 8416
            Fax: 301 504 8414
            Email: tads@psi.barc.usda.gov
            Single pass sequencing. Bases called and alt_trimmed with phred

```

v0.980904.e. Vector identified by cross_match with the -minscore 18 and -timatch 12 options.

PCR Primers

FORWARD: AGGAACAGCTATGACCAT
 BACKWARD: GTTTCACAGTCACGACG
 Plate: 13 row: C column: 17
 Seq primer: ATTTAGCTGACACTATAG.

FEATURES

source

1.490

/organism="Bos taurus"

/db_xref="taxon:9913"

/clone_lib="BARC 5BOV"

/tissue_type="pooled"

/lab_host="DH10B"

/note="Vector: pCMV SPORT6; Site_1: XbaI; Site_2: XhoI; library made from pooled mRNA isolated from mammary tissues at eight physiological, developmental, and disease states."

Location/Qualifiers

1.490

BASE COUNT 124 a 128 c 145 g 93 t

ORIGIN

Query Match 21.0%; Score 365.2; DB 136; Length 490;
 Best Local Similarity 84.1%; Pred. No. 1.4e-94;
 Matches 412; Conservative 0; Mismatches 78; Indels 0; Gaps 0;

QY 1137 gaacaatgacattgagctatgaagctgcaagaagctctgacttaacgactatgaa 1196

Db 1 GAACAAATGACATTTGCGCTATGAACTGACAGAGCTCTGACTTCAACGACAAAGTAA 60

QY 1197 accagtgctcgcgcaaccagcagcagcagcagcagcagcagcagcagcagcagc 1256

Db 61 GCAGGTGTGCTACCAATACAGGCAATGATGCTGAGGCAACACACATCTCCGTGATTTTC 120

QY 1257 cgggtgaggggccacccagcagcagcagcagcagcagcagcagcagcagcagc 1316

Db 121 CGGCTGGGGGGCCACCTACGAGAAAGGAAAGACCTCAGATGACCTGACCGGCCAAGGT 180

QY 1317 gcttcctatgagacacagagatgacacagcagcagcagcagcagcagcagcagc 1376

Db 181 GCACCTCATTTGACCCCGGAGAGTGCACAGCAAGTACATGTTAGACAACTCATACGCC 240

QY 1377 agcagatgctcgcgctcgcgctcgcgctcgcgctcgcgctcgcgctcgcgctcgc 1436

Db 241 CGCAGATGCTGACGAGCTACTGCGGGGCAACCGTCACTCTGCGAGGCTGACAGTGG 300

QY 1437 agggctcgtgctcactcgaagaacaatactggtgctgataggagatacaagctgg 1496

Db 301 AGGCTCTCTGCTACACTGAAAGACAGTGTCTGCTGCTGATTTGAGACACGAGCTGGGG 360

QY 1497 ttctgctctgcaaaagcttacagcagcagcagcagcagcagcagcagcagcagcagc 1556

Db 361 GTCAGGCTGTGCGCAAGCGTACCGACCGGAGGTACGGGAACTGACAGTGTTCACAGA 420

QY 1557 ctggaattatcgacaatagaggcagcagcagcagcagcagcagcagcagcagcagcagc 1616

Db 421 CGGAGCTGATCAACAATAGGGGCAAAAGCTATCTGTATGGCGTGGGTCTTGACAAAC 480

QY 1617 gctttacaag 1626

Db 481 ATTGCACAAG 490

RESULT 14

LOCUS

BF786954

DEFINITION

602112148P1 NCI_CGAP_Kid14 Mus musculus cDNA clone IMAGE:4240378

ACCESSION

BF786954

VERSION

BF786954.1

KEYWORDS

EST.

SOURCE

ORGANISM

Mus musculus

REFERENCE

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 788)
 NIH-MGC <http://mgc.nci.nih.gov/>.
 TITLE
 JOURNAL
 COMMENT
 Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-remail.nih.gov
 Tissue Procurement: Jeffrey E. Green, M.D.
 cDNA Library Preparation: Life Technologies, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNLN at: <http://image.llnl.gov>
 Plate: LLM9854 row: p column: 11
 High quality sequence stop: 662.

FEATURES

source

1.788

/organism="Mus musculus"

/strain="FVB/N"

/db_xref="taxon:10090"

/clone_lib="IMAGE:4240378"

/clone_lib="NCI_CGAP_Kid14"

/lab_host="DH10B (T1 phage-resistant)"

/note="Organ: Kidney; Vector: pCMV-SPORT6; Site_1: NotI; Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT. Average insert size 1.75 kb. Constructed by Life Technologies. Note: this is a NCI_CGAP Library."

Location/Qualifiers

1.788

BASE COUNT 198 a 198 c 207 g 185 t

ORIGIN

Query Match 20.5%; Score 355.8; DB 169; Length 788;
 Best Local Similarity 80.5%; Pred. No. 8.9e-92;
 Matches 453; Conservative 0; Mismatches 107; Indels 3; Gaps 3;

QY 1067 tcatgttctatgagccgagatgacagcagcagcagcagcagcagcagcagcagcagcagc 1126

Db 13 TCATGTTCTATGAGAGTAGACACCGTACGAAAGTAATTTCCATCCAAATTAGACT 72

QY 1127 ccaagaccgaagaacaatgacatgctgctgataagcagcagcagcagcagcagcagcagc 1186

Db 73 CTAAAGCAAGAAATTAACGACATGCTCTCATGAGCTGACGACACCT-TGCTTTAAAG 131

QY 1187 acctagtgaaccagctgctcgcgcaaccagcagcagcagcagcagcagcagcagcagc 1246

Db 132 ATCTAGTGAAGCCAGTGTCTGCTCCGAAACCAAGCATGATGACTAGACCTGACAGGAAT 191

QY 1247 gctggaattccgggtgaggggccacccagcagcagcagcagcagcagcagcagcagcagcagc 1306

Db 192 GCTGGATTTTGGGGGTGGGGGCCACCTATGAGAAAGGAGCTGGAGGTGTAATG 251

QY 1307 ctgccaagctgctcctcatgagaacacagatgacacagcagcagcagcagcagcagcagcagc 1366

Db 252 CTGCGATGTGATCCCTGATGAGCCCTCCAAATGTATATGTAATATACATACAAAC 311

QY 1367 tgatcaaccagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1426

Db 312 TAATTCACACACGATGATCTGTGCGGCTCTCCACAGGGGTCTGTCGACTCTTGCA 370

QY 1427 gtgcaagtgaggagcctctgctcactcgaagaacaatactggtgctgataggagata 1486

Db 371 GAGACAGTGGAGGGCGCGCTGTGTTTGAAGAAATGGAGATCTGTGGCTGATTTGGGAC 430

QY 1487 caagctgggggtctgctgctgccaagcctacagacagcagcagcagcagcagcagcagcagc 1546

Db 431 CGAGCTGGGGCTCGGGCTGTGCCAAGGCACTCAGACCTGAGATACGGGAAAGTGACG 490

QY 1547 taatcaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1606

Db 491 TATTACAGATTGATCTACACCAATAGAGGCGAAGCACTAATCCAGCTGG-CTTTGT 549

QY 1607 ccttgacgctgtttaccagaaga 1629
 Db 550 CCCAGACTTCCTTGTCTCAAA 572

RESULT 15
 LOCUS BG242090 910 bp mRNA EST 13-FEB-2001
 DEFINITION 602354634F1 NCI_CGAP_Mam1 Mus musculus cDNA clone IMAGE:4482777 5',
 mRNA sequence.
 ACCESSION BG242090
 VERSION BG242090.1 GI:12751905
 KEYWORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 910)
 NIH-MGC http://mgc.nci.nih.gov/.
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgepds-remail.nih.gov
 Tissue Procurement: Gilbert Smith, Ph.D.
 cDNA Library Preparation: Life Technologies, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 Plate: LLM10320 row: d column: 10
 High quality sequence stop: 644.
 Location/Qualifiers
 1. 910
 /organism="Mus musculus"
 /strain="FVB/N"
 /db_xref="taxon:10090"
 /clone_1b="IMAGE:4482777"
 /clone_1lb="NCI_CGAP_Mam1"
 /tissue_type="tumor, biopsy sample"
 /dev_stage="3 months, virgin"
 /lab_host="DH10B"
 /note="Organ: mammary; Vector: pCMV-SPORT6; Site_1: SalI;
 Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT.
 Library constructed by Life Technologies. Investigator
 providing samples: Gilbert Smith, NIH"

BASE COUNT 206 a 236 c 238 g 230 t
 ORIGIN

Query Match 19.5% Score 338.4: DB 175: Length 910;
 Best Local Similarity 75.0% Pred. NO. 1e-86;
 Matches 555: Conservative 0: Mismatches 171: Indels 14: Gaps 10:

QY 339 gtgcacccaagaactaagaacacgtgcatcaacctgacccctggagaccccttcctcgt 398
 Db 1 GTGCACCTCAAGACTAAGAAATCGTGTAGCCCTTGCCCTGGGCACCTGCTCTCAC 60

QY 399 gggagactgcgcgtggcgcgtgactactcgtgaagtcatggcgagcaagtgctcaactc 458
 Db 61 GGGAGCTGCTGCTGCTGCTGCTTTGGAGGTTCTGGACAGCACTGTTCTACGTC 120

QY 459 tgggataagatgcagactcctcagtaactcaaaccccttaactggatgtagtgcgt 518
 Db 121 TGAGATGAGGTGTGGGTCTTCAGGCACATCATCACTTCTCTGTGTGAGCGGT 180

QY 519 gtcaacacgcccggcg- ggaagaagaaatcggtgttgcctctaaggacaact 577
 Db 181 AGCACATTGTGCCAAGCGGAGAGATGAGAACGTTGTGTTGCTCTACGAGCAAAAGCT 240

QY 578 tcatctcaggtgtaactcatctcagaggaagtcctgcaaccctgtgtgccaagagact 637
 Db 241 TCATCTCCAGGTTTACTCATCTCAGAGAAAGCCTGGTATCCCGTGTGCCAGATGATT 300

QY 638 ggaacgagaactacgagcgcgctgcaaggagacatggcctataagaataatttact 697
 Db 301 GGAGTGAGAGCTACAGCGGAGAGACGACGATGTAAAGACATGGGATACAAAGAACTTTTATT 360

QY 698 ctaccgaagaatagtgatgacagcggaatccacgaactta-tgaactgaacacaagt 756
 Db 361 CTACGCCAAGGATACACAGACGAGCGGGGCAACGACTTTACTGAAGCTGAATGTGAGC 420

QY 757 gccggcaatgcatatcat-aaaaactgacacagatgagctgttcttcaaaagc 815
 Db 421 TCAGGCAATGTGACCTCTATATAAAACTCTACCAACGTGACTATGTATCCGCAT 480

QY 816 agtgg-ttctttaacgtgatacctgctggggtcaacttgaactcaaacgcgcagagca 874
 Db 481 GGTGGTTTCTTGGCGCTGTATAGAAATGGGGTTCGCT---CACTGAAGCCGACAGACA 537

QY 875 ggaatgtggcgcgagagcg-cgctccgggggcttgccctggcaggtcagctcgaac 933
 Db 538 GGATTGTGGGTGATTTGAATGCTCACCAGGAGACTGGGCTCGCAGGTCCCTGCAC 597

QY 934 gtccagaagctcaagtgtaggaagct-catcataccccccagtgatcgtgacagc 992
 Db 598 GTCAAGCGCTCACGCTGCGAGGCTCCATCATCACCCGAGTGGATTTGTACCGG 657

QY 993 cgccactgcgtggaaaaacctttaacaatcatgacatgtgaagcgaattgcgggagat 1052
 Db 658 CG-CGATTGTGTGGAAGACCC---AGGGGCGGAGGTACTGAGCCGCAATTTC-GGAAT 712

QY 1053 tttagacaactcttcacatgt 1072
 Db 713 TCTGGACACAGTTTCCGTGT 732

Search completed: September 26, 2001, 11:02:02
 Job time: 6029 sec

THIS PAGE BLANK (USPTO)

Dh 301 agaaagcactgtgcatcaaccctgagcccttgaggacctctcgttgaggactgcgtgcg 360
Oy 416 cggcgactctcggaaatctcatcggcagcaagtgtcacaactctgggacagagtgcact 475
Dh 361 cggcgactctcggaaatctcatcggcagcaagtgtcacaactctgggacagagtgcact 420
Oy 476 cctcaggtacatcgaacaccccttaactggtgagtcgtgacgtgtcacaactcggcg 535
Dh 421 cctcaggtacatcgaacaccccttaactggtgagtcgtgacgtgtcacaactcggcg 480
Oy 536 gggaggagcagaaatcgggtgtgttgcctctacagcaaaccttcaactccttcgggtact 595
Dh 481 gggaggagcagaaatcgggtgtgttgcctctacagcaaaccttcaactccttcgggtact 540
Oy 596 catccagaggaagtcctctgcaaccctgtgtgcacagacagctggaacggaactaaggcc 655
Dh 541 catccagaggaagtcctctgcaaccctgtgtgcacagacagctggaacggaactaaggcc 600
Oy 656 gggcgactcgcagggagcattggtctataagaataatttctaactctagcaaggatagtg 715
Dh 601 gggcgactcgcagggagcattggtctataagaataatttctaactctagcaaggatagtg 660
Oy 716 atgacagcggatccacacagcttataagaataatttctaactctagcaaggatagtg 775
Dh 661 atgacagcggatccacacagcttataagaataatttctaactctagcaaggatagtg 720
Oy 776 ataaaaactctgacacagctgacgtgtctctcaaaaagcagtggtcttctacgcgtga 835
Dh 721 ataaaaactctgacacagctgacgtgtctctcaaaaagcagtggtcttctacgcgtga 780
Oy 836 tagctctcggtgtcaacttgaactcaagccgcagagcagagatgtggcgcgagagcg 895
Dh 781 tagctctcggtgtcaacttgaactcaagccgcagagcagagatgtggcgcgagagcg 840
Oy 896 cgtcccgaggcgctgggacctggcaggtcaggtcagctcagaaagctccagctgtgcg 955
Dh 841 cgtcccgaggcgctgggacctggcaggtcaggtcagctcagaaagctccagctgtgcg 900
Oy 956 gaggctccatcaccaccccgagtgatcgtgacagcgccacatgtcgtggaacacctc 1015
Dh 901 gaggctccatcaccaccccgagtgatcgtgacagcgccacatgtcgtggaacacctc 960
Oy 1016 ttaacaatcattggaatctggacggcatttgcgggatttctggaacattcttcagtctt 1075
Dh 961 ttaacaatcattggaatctggacggcatttgcgggatttctggaacattcttcagtctt 1020
Oy 1076 atgagcgcgagacacaaagtagaattctcatcaaatatgactccaagacca 1135
Dh 1021 atgagcgcgagacacaaagtagaattctcatcaaatatgactccaagacca 1080
Oy 1136 agaaacatgacatctgcgtctgataagctgtcagaagcctctgaacttcaacgactagta 1195
Dh 1081 agaaacatgacatctgcgtctgataagctgtcagaagcctctgaacttcaacgactagta 1140
Oy 1196 aaccagtggtctgcacacacacagcagatgactgcagcacaacagcctctctgattt 1255
Dh 1141 aaccagtggtctgcacacacacagcagatgactgcagcacaacagcctctctgattt 1200
Oy 1256 cggggtgaggggcagcagcagaggaaggaagcctcagaagtgtcgaacgctctgcaag 1315
Dh 1201 cggggtgaggggcagcagcagaggaaggaagcctcagaagtgtcgaacgctctgcaag 1260
Oy 1316 tgcctctcatctgagacacagagatgcaacagcagatagtgctatagaacacgtatcac 1375
Dh 1261 tgcctctcatctgagacacagagatgcaacagcagatagtgctatagaacacgtatcac 1320
Oy 1376 cagcagatgactgtgcggctcctcgcagggaggaagtcgacttgcgcagggtgacagtg 1435
Dh 1321 cagcagatgactgtgcggctcctcgcagggaggaagtcgacttgcgcagggtgacagtg 1380
Oy 1436 gaggcctctgtcactcctcagaaacaatatctgtgtgcgtgataaggagatacaagctggg 1495
Dh 1381 gaggcctctgtcactcctcagaaacaatatctgtgtgcgtgataaggagatacaagctggg 1440

Oy 1496 gtctcgtctgtccaaagcttacagacagaggtgtacgggaatgtgatgtatccag 1555
Dh 1441 gtctcgtctgtccaaagcttacagacagaggtgtacgggaatgtgatgtatccag 1500
Oy 1556 actgattatctgacaaatgagggcagagcgtcaatccaatgtctctcgttgaagt 1615
Dh 1501 actgattatctgacaaatgagggcagagcgtcaatccaatgtctctcgttgaagt 1560
Oy 1616 cgtttacagaagaacaatgagggtgtgtgtcttcccggtcagtattactcttagag 1675
Dh 1561 cgtttacagaagaacaatgagggtgtgtgtcttcccggtcagtattactcttagag 1620
Oy 1676 atgattcagaaggtcacttcatcttattatcaacagtgaactgtctgcg 1723
Dh 1621 atgattcagaaggtcacttcatcttattatcaacagtgaactgtctgcg 1668

RESULT 2
US-09-342-749-1
: Sequence 1, Application US/09342749
: Patent No. 6166194
: GENERAL INFORMATION:
: APPLICANT: Mong, Alexander K.C.
: APPLICANT: Tavligian, Sean V.
: APPLICANT: Teng, David H.-F.
: APPLICANT: Myriad Genetics, Inc.
: TITLE OF INVENTION: TPRS2 is a Tumor Suppressor
: FILE REFERENCE: 2318-202
: CURRENT APPLICATION NUMBER: US/09/342,749
: EARLIER FILING DATE: 1999-06-29
: EARLIER APPLICATION NUMBER: US 60/091,044
: NUMBER OF SEQ ID NOS: 33
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 1
: LENGTH: 1479
: TYPE: DNA
: ORGANISM: Homo sapiens
: FEATURE:
: NAME/KEY: CDS
: LOCATION: (1)..(1476)
: FEATURE:
: NAME/KEY: conflict
: LOCATION: (724)
: OTHER INFORMATION: Listed as T in GenBank Accession NO. U75329
: FEATURE:
: NAME/KEY: conflict
: LOCATION: (985)
: OTHER INFORMATION: Listed as C in GenBank Accession No. 6166194 U75329
: FEATURE:
: NAME/KEY: conflict
: LOCATION: (1347)
: OTHER INFORMATION: Listed as C in GenBank Accession No. 6166194 U75329
: FEATURE:
: NAME/KEY: conflict
: LOCATION: (1466)
: OTHER INFORMATION: Listed as A in GenBank Accession No. 6166194 U75329
: FEATURE:
: NAME/KEY: conflict
: LOCATION: (1471)
: OTHER INFORMATION: Listed as A in GenBank Accession No. 6166194 U75329.
: FEATURE:
: NAME/KEY: allele
: LOCATION: (478)
: OTHER INFORMATION: This base can be G or A with G being the more
: OTHER INFORMATION: common allele. The codon will change from Val to
: OTHER INFORMATION: Met.
: FEATURE:
: NAME/KEY: allele
: LOCATION: (777)
: OTHER INFORMATION: This base can be C or T with C being the more
: OTHER INFORMATION: common allele. The codon is unaffected with both

OTHER INFORMATION: alleles encoding Gly.
FEATURE:
NAME/KEY: allele
LOCATION: (768)
OTHER INFORMATION: This base can be C or T with C being the more
OTHER INFORMATION: common allele. This is a silent polymorphism.
FEATURE:
NAME/KEY: allele
LOCATION: (834)
OTHER INFORMATION: This base can be C or T with C being the more
OTHER INFORMATION: common allele. This is a silent polymorphism.
FEATURE:
NAME/KEY: allele
LOCATION: (625)
OTHER INFORMATION: This base can be T or A with T being the more
OTHER INFORMATION: common allele. The codon will change from Phe to
OTHER INFORMATION: Ile
US-09-342-749-1

Query Match 85.0%; Score 1477.4; DB 4; Length 1479;
Best Local Similarity 99.9%; Pred.No. 0;
Matches 1478; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 112 atggtcttgaactcaaggtcaccacagctatgtgacctactatgaataacatgatac 171
DB 1 atggtcttgaactcaaggtcaccacagctatgtgacctactatgaataacatgatac 60
QY 172 caacgggaaacccctatcccgacagcccaatgtgtcccaatgtctacgaagtgcac 231
DB 61 caacgggaaacccctatcccgacagcccaatgtgtcccaatgtctacgaagtgcac 120
QY 232 ccgggtcactacacccgtcccccgtcccccagtaagcccccgggttcctgaagccgct 231
DB 121 ccgggtcactacacccgtcccccgtcccccagtaagcccccgggttcctgaagccgct 180
QY 292 tccaaccccgctcgtcgcacgacgcccacatcccccacgagacgtgtgcacctaaag 351
DB 181 tccaaccccgctcgtcgcacgacgcccacatcccccacgagacgtgtgcacctaaag 240
QY 352 actaagaagacactgtgtacacacacacacacacacacacacacacacacacacac 411
DB 241 actaagaagacactgtgtacacacacacacacacacacacacacacacacacacac 300
QY 412 gcgcgtgcctactcctggaagttcatgagggcaggaagtgctcccaactcgtgataagtg 471
DB 301 gcgcgtgcctactcctggaagttcatgagggcaggaagtgctcccaactcgtgataagtg 360
QY 472 gactcctcagtaactcgtacacacacacacacacacacacacacacacacacacac 531
DB 361 gactcctcagtaactcgtacacacacacacacacacacacacacacacacacacac 420
QY 532 ggcgggaggaagacagaatcgggtgtgtcgccttacgacacacacacacacacacac 591
DB 421 ggcgggaggaagacagaatcgggtgtgtcgccttacgacacacacacacacacacac 480
QY 592 tactcatcagaggaaggtcctcgtgacacacacacacacacacacacacacacacac 651
DB 481 tactcatcagaggaaggtcctcgtgacacacacacacacacacacacacacacacac 540
QY 652 ggcgggaggaagacagaatcgggtgtgtcgccttacgacacacacacacacacacac 711
DB 541 ggcgggaggaagacagaatcgggtgtgtcgccttacgacacacacacacacacacac 600
QY 712 gttgagtagacagcgatccacacacacacacacacacacacacacacacacacac 771
DB 601 gttgagtagacagcgatccacacacacacacacacacacacacacacacacacac 660
QY 772 atctataaaaactgtac 831
DB 661 atctataaaaactgtac 720
QY 832 tgtatagctgcgggggtcaactgtgaactcaagccgacagacagatgtgtgcccggag 891

DB 721 tgtatagctgcgggggtcaactgtgaactcaagccgacagacagatgtgtgcccggag 780
QY 892 agcgctcctccgggggctgtgcccctgacaggtcagctcagctcagctcagctcagct 951
DB 781 agcgctcctccgggggctgtgcccctgacaggtcagctcagctcagctcagctcagct 840
QY 952 tgcggaggtcctac 1011
DB 841 tgcggaggtcctac 900
QY 1012 cctcttaacacatcaggtcattgtgacgacattgtgcgggatttttgagacatcttcag 1071
DB 901 cctcttaacacatcaggtcattgtgacgacattgtgcgggatttttgagacatcttcag 960
QY 1072 ttctatgagcgagatacagaatagaaaagtgtatttctcattcacaattatgactccag 1131
DB 961 ttctatgagcgagatacagaatagaaaagtgtatttctcattcacaattatgactccag 1020
QY 1132 accaagaacatgacattgtgctgataagcttcagaaagcttcgacttcacacagccta 1191
DB 1021 accaagaacatgacattgtgctgataagcttcagaaagcttcgacttcacacagccta 1080
QY 1192 gtgaacacagtgctgtcccaacccagagatgctgtgacgacgaagacagctctgtg 1251
DB 1081 gtgaacacagtgctgtcccaacccagagatgctgtgacgacgaagacagctctgtg 1140
QY 1252 attccgggttggggggccacggaggaaggaaggaaggaaggaaggaaggaaggaag 1311
DB 1141 attccgggttggggggccacggaggaaggaaggaaggaaggaaggaaggaaggaag 1200
QY 1312 aaggtctctcatttgaagacagagatgtgaacagagatgtctatgacacacccgatac 1371
DB 1201 aaggtctctcatttgaagacagagatgtgaacagagatgtctatgacacacccgatac 1260
QY 1372 ac 1431
DB 1261 ac 1320
QY 1432 agtggagggcctcgtgac 1491
DB 1321 agtggagggcctcgtgac 1380
QY 1492 tgggttctcgtgctgcac 1551
DB 1381 tgggttctcgtgctgcac 1440
QY 1552 acgagctgattatcgac 1590
DB 1441 acgagctgattatcgac 1479

RESULT 3
US-08-807-151-2
Sequence 2, Application US/08807151
Patent No. 6043033
GENERAL INFORMATION:
APPLICANT: Bandman, Olga
APPLICANT: Lal, Preeti
TITLE OF INVENTION: NOVEL HUMAN PROSTATE-ASSOCIATED
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:
ADDRESSER: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: US
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS

```
SOFTWARE: FASTSEQ Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/807.151
FILING DATE: Filed Herewith
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0227 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-855-0555
TELEFAX: 415-845-4166
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1077 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: SCORNO701
CLONE: 556016
US-08-807-151-2
```

```
Query Match          57.1% Score 992.4; DB 3: Length 1077;
Best Local Similarity 99.8% Pred. No. 1.6e-260;
Matches 993; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```

```
OY 721 agcgagatccacagcttctgaactgaacacaaagtcgagcaatgcatctataaa 780
DB 83 accgagatccacagcttctgaactgaacacaaagtcgagcaatgcatctataaa 142
OY 781 aacagtgaccacagagatgctgcttctcaaaagtgcttcttaagctgataagc 840
DB 143 AACGTGTACACAGTGTATGCTTTCAAAAGCAGTGTCTTTACGCTGATATAGCC 202
OY 841 tgcggggtcaacttgaactcaagccgacagagagatctgaggcgagagcgagc 900
DB 203 tgcggggtcaacttgaactcaagccgacagagagatctgaggcgagagcgagc 262
OY 901 ccgaggggctgagccttgagagtcagcctgcaagtcagagacgctcagtgagagc 960
DB 263 CCGGGGGCTTGCGCTGCGAGGTCAAGCTGTCAGAACGTCGACGTCGAGGAGC 322
OY 961 tccatcatcaccgagagtgatgctgtagacgagccgacgctgctgtaaaactctaac 1020
DB 323 tccatcatcaccgagagtgatgctgtagacgagccgacgctgctgtaaaactctaac 382
OY 1021 aatcgaatgcatctgagagcatttgcgggagcttgaagacaactcttcaatgta 1080
DB 383 aatcgaatgcatctgagagcatttgcgggagcttgaagacaactcttcaatgta 442
OY 1081 gccgagatcacaagtgagaaagtgatcttcaatccaaltatgactccaagaacaagc 1140
DB 443 GCCGGATACCAAGTGAAGAAAGTGAATTTCTCATCCAAATTAAGACTCCAAGCAAGAAC 502
OY 1141 aatgacattgagctgaatgagtgtagaagccttgcagcttcaagagcaatgtaaa 1200
DB 503 aatgacattgagctgaatgagtgtagaagccttgcagcttcaagagcaatgtaaa 562
OY 1201 gctgtctgcccacacagagatgctgtagcagcagaacagctctgctggaattccgg 1260
DB 563 gctgtctgcccacacagagatgctgtagcagcagaacagctctgctggaattccgg 622
OY 1261 tggggggccacagagagagaaagagacctcaagaatgctgtaagcctgccaagtgctt 1320
DB 623 TGGGGGGCCACCGAGAGAGAAAGGAAAGTCAAGAGTCTGAACGCTGCCAAGGTGCTT 682
OY 1321 ctcatgtgagacacagagatgcaacagcgatattgtctatagacaactgtatcacagc 1380
|||||
```

```
DB 683 CTCATTGAGACACAGAGATGACAGACAGATATGCTTATGACACACCTGATCACACGCC 742
OY 1381 atgactgtgccgctctcctgcaaggaaacgctcgaattcttccagagtgacagtgaggg 1440
DB 743 ATGATCTGTGCCGGCTCTCTGACGGGAACGTCGATCTTGGCAGGGTACAGTGGAGGG 802
OY 1441 cctctgttacttgcgaagaacatctggtgctgataaggagataaagcctggagctt 1500
DB 803 CMTGTGTCACCTGCAAGAACATATCTGTGGCTGATAGGGGATACAAAGCTGGGGTCT 862
OY 1501 ggcctgtgccaaagcttaagcagagagtgtagggaaatgtagtatttccagcagcgg 1560
DB 863 GCGTGTGCCAAAGCTTACAGACAGAGAGTACGGGAATGTATGTATTCACGAGACTGG 922
OY 1561 attatcgaacaaatgagagcagagcctaatacgaatgctcgtcctgagcgtctt 1620
DB 923 ATTTATCGACAAATAGAGGCAACGCTTAATCAGATGCTCTTCCCTTGTGACGCTTT 982
OY 1621 tacaagaacaaatgggagctgttcttcccgctgcatgattactcttagagatgat 1680
DB 983 TACAAGAAACAAATGGGCGCTGTTTGTCTCCCGTCATGATTTACTTATGAGATGAT 1042
OY 1681 tcagagtgactctatttatttaaacagtgact 1715
DB 1043 TCAGAGGTCATCTTATTTATTAACAGTGAAT 1077
```

```
RESULT 4
US-09-008-271A-18
Sequence 18, Application US/09008271A
Patent No. 6203979
GENERAL INFORMATION:
APPLICANT: Bandman, Olga
Hillman, Jennifer L.
Yue, Henry
Guegler, Karl J.
Corley, Neil C.
Tang, Tom Y.
Shah, Purvi
TITLE OF INVENTION: HUMAN PROTEASE MOLECULES
NUMBER OF SEQUENCES: 24
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Dr.
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/008,271A
FILING DATE: 16-Jan-1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: <Unknown>
FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Mohan-Peterson, Sheela
REGISTRATION NUMBER: 41,201
REFERENCE/DOCKET NUMBER: PF-0458 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-855-0555
TELEFAX: 650-845-4166
INFORMATION FOR SEQ ID NO: 18:
SEQUENCE CHARACTERISTICS:
LENGTH: 2038 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
```



```

Db 574 TTATATGGGACCCACCTGTGTGGGGGTCCTGCTGTCTGGGACTGGGTCTGATGTC 633
Qy 993 cggccacgtcgctggaagaaacccctcttaacatcatgacatgagacatctgacggagat 1052
   || || || || || || || || || || || || || || || || || || || || || ||
Db 634 TGCAATATGCTTTCCAGAGCGGAAACCGGGTCTGTCTGCTGGTGGCAGATTATTGCTGTGC 693
Qy 1053 ttctgagacaatcttca-----tctctatcgagccggatctacaaatgaa 1097
   || || || || || || || || || || || || || || || || || || || || || ||
Db 694 TGTACCCCGGACCTCAACCCCATGCTGTGCACACTGGGGGCTTACGCTGTATCATATG 753
Qy 1098 aaaaagtattctctatccaaatctatgactccaaagccaaagaaatgcatcgctgat 1157
   || || || || || || || || || || || || || || || || || || || || || ||
Db 754 GGGCTACCTCTCTTTCGAGACCTTACTATGAGCAAAACACAGATGACATTGCTTGCT 813
Qy 1158 gaagctgcagaagacccctctgactctcaacgaacctagtgaaacagatgctgctccaaacc 1217
   || || || || || || || || || || || || || || || || || || || || || ||
Db 814 CCACCTCTCTAGCTCCCTGCGCTCTCAGAAATCATTCACAGCCAGTGTGTCTCCCTGTGTC 873
Qy 1218 aggcatagtctgcagccagaaagctctgctggaatctccggatggggggccaccgaaaga 1277
   || || || || || || || || || || || || || || || || || || || || || ||
Db 874 GGGACAGGCGCTGTGTGATGGCAGAGCTGTGTGACCGGCTGGGTAAACACACAGTT 933
Qy 1278 gaagaggaaagacctgaagaagctgcgaagctgcgaagtgctctcatgagacaagag 1337
   || || || || || || || || || || || || || || || || || || || || || ||
Db 934 CTATGGCCACAGGCTTATGTGTCTCCAGAGGCCGGGTTCCCATCATTAAGCAACAGACT 993
Qy 1338 atgcacaacagagatagtctatgacaacctgataccaaacagcatgactgctgcggctt 1397
   || || || || || || || || || || || || || || || || || || || || || ||
Db 994 TTGCACACAGCCCGGACTTCTAGGGAAATCATGATCAAGCCCAAGTGTCTGTGCTGAGTA 1053
Qy 1398 cctgcagagggaagctcgactctctgcacaggtgacagtggaagggcctctggt----- 1448
   || || || || || || || || || || || || || || || || || || || || || ||
Db 1054 TCTGTAGGCTGCAATGATGCTGTCAGAGGCGACAGATGAGAGCCCTTGTGTGAAGA 1113
Qy 1449 ---cacttcgaagaaacaatatctgtgtgctgataagggaatacaagctggtctcgctg 1505
   || || || || || || || || || || || || || || || || || || || || || ||
Db 1114 CAGCATCTCTGGGACATCAAGATGGCGGCTATGTGCAATTGAACTGGGGATACGGGCTG 1173
Qy 1506 tgcacaagcttaccagacaggaagtacaggaatgatagtatcatcagcagactgattta 1565
   || || || || || || || || || || || || || || || || || || || || || ||
Db 1174 TGCTTTGGCCCGGACGAGCAGAGTGTACCAACAACTCACTGCCGAGCTGATCTT 1233
Qy 1566 tcgacaatatgagggcagacggttaacca 1594
   || || || || || || || || || || || || || || || || || || || || || ||
Db 1234 CAAGGCATTAAGACTCACTCCGAGGCCA 1262

```

RESULT 6
US-09-510-738A-188
; Sequence 188, Application US/09510738A
; Patent No. 6268165
; GENERAL INFORMATION:
; APPLICANT: O'Brien, Timothy J.
; TITLE OF INVENTION: Compositions and Methods for the Early Diagnosis of
; FILE REFERENCE: D6233CIP-A
; CURRENT APPLICATION NUMBER: US/09/510,738A
; PRIOR FILING DATE: 2000-02-22
; PRIOR APPLICATION NUMBER: 09/039,211
; NUMBER OF SEQ ID NOS: 188
; SEQ ID NO 188
; LENGTH: 1783
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: full length cDNA of hepsin
US-09-510-738A-188

Query Match 7.1%; Score 124.2; DB 4; Length 1783;
Best Local Similarity 51.3%; Pred. NO. 2.3e-24;
Matches 384; Conservative 0; Mismatches 330; Indels 27; Gaps 3;

```

Qy 873 caggaatctgggcggcgagagcgctcccgggggcccttgccttgcaggtcagcttca 932
   || || || || || || || || || || || || || || || || || || || || || ||
Db 728 ccgacatctgaggagcgcgagacacagcttggcgcggtgagccggtgcaagtacagcttcg 787
Qy 933 cgtccagaacgtccacgtgtgcgagagctccatcatcaaccccgagtgatctgcagc 992
   || || || || || || || || || || || || || || || || || || || || || ||
Db 788 ctatgatgagacacactctgtggggaatccctgctctcccgaggacttgggtgtctgacgc 847
Qy 993 cggccacgtcgctggaagaaaccccttaacaatccatgagatggatggaaatttgcgggag 1052
   || || || || || || || || || || || || || || || || || || || || || ||
Db 848 cggccacgtctccggagcggaacgggtctctgtcccgatggcgagtggttgcgtgtgc 907
Qy 1053 ttctgacaatcttcatgtctat-----ggagccgatatcaagatgaagaaagt 1103
   || || || || || || || || || || || || || || || || || || || || || ||
Db 908 cgtggcccaagcctctccccaagctctgacgtgggggtgagcgtgtggtctacacag 967
Qy 1104 gattctcatcaaatatgatctccaagaacaa-----caatgaatctgctgat 1157
   || || || || || || || || || || || || || || || || || || || || || ||
Db 968 gggctatcttccctcttcggaccccaacagcgaagagaacagcaacgatatgtccctgt 1027
Qy 1158 gaagctgcagaagcctctgacttcaacgaacctagtgaaacagatgctctgcccacacc 1217
   || || || || || || || || || || || || || || || || || || || || || ||
Db 1028 caactctccagctccctccgcccctacagaatacatccagcctgtgtgtctccagctgc 1087
Qy 1218 aggcatagtctgcagccagaacagctctgtggaattccgggttggggccaccgaaaga 1277
   || || || || || || || || || || || || || || || || || || || || || ||
Db 1088 cggccagcgccctgtgtgatagtgcaagatctgtacgctgacggcggtgggcaacagcagta 1147
Qy 1278 gaagaggaaagacctgaagaagctgtgaacgctgacaaagtgtctctcatatggacaaga 1337
   || || || || || || || || || || || || || || || || || || || || || ||
Db 1148 ctatgccaacaagcgcgggggtatctccaggaagctctgaagttcccaatacaaatgaatgt 1207
Qy 1338 atgcacaacagagatagtctatgacaacctgataccacacagcatgactgctgcggctt 1397
   || || || || || || || || || || || || || || || || || || || || || ||
Db 1208 ctgcacatgctgtgactctctatgaaacagatacaagccaagaatgtctgttcggctta 1267
Qy 1398 cctgcaggggaagctcgactctctgcacaggtgacagtggaagggcctctggt----- 1448
   || || || || || || || || || || || || || || || || || || || || || ||
Db 1268 ccccgagggtgtgcatgtgctgcacagggcgacagcggtgttcccttgtgtgaaaga 1327
Qy 1449 ---cacttcgaagaaacaatatctgtgtgctgatagggaatacaagcttgggtctgctg 1505
   || || || || || || || || || || || || || || || || || || || || || ||
Db 1328 cagcatctcttcgagcgcaacgttggcggtctgtgtgcatgtgagttvggcaacagctgt 1387
Qy 1506 tgcacaagcttaccagacaggaagtacaggaatgatagtatcatcagcagctgattta 1565
   || || || || || || || || || || || || || || || || || || || || || ||
Db 1388 tgccttggccagaaagcagcagcgtctacacaaagtcatgactcgggaagtgtactt 1447
Qy 1566 tcgacaatatgagggcagacggttaacca 1594
   || || || || || || || || || || || || || || || || || || || || || ||
Db 1448 ccaggcataaagactcaactccgaagca 1476

```

RESULT 7
US-08-200-900A-1
; Sequence 1, Application US/08200900A
; Patent No. 5665566
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: CLONING OF ENTEROKINASE AND METHOD OF USE
; NUMBER OF SEQUENCES: 38
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc. - Legal Affairs
; STREET: 87 Cambridgepark Drive
; CITY: Cambridge
; STATE: MA
; COUNTRY: USA
; ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS


```

Db 420 AATAAGTAATGATGTAATGTAATGACACCATAGTATTAATGAGCCATCTTGTCTGGAAT 479
Oy 1383 gattctgctggcgtctccctcgaagggaagtcgactcttgcgaaggtgagacagtgaggcc 1442
Db 480 GCTGTGTGCTGGAGTACCTCAAGGTGAGTGTGAGCATGTCAAGGGTACTGTGTGGCCCC 539
Oy 1443 tctgtgactctgaa--gaacaatactgtgtgctgataagggatatacaagcttggtctc 1499
Db 540 ACTAGTACAAGAAAGACTACGCGCGCTTGTATTATTTGTGGGATAGTAAGCTGGGGAGA 599
Oy 1500 tggcgtgtcgaagcttacaagccaggagtgtaacgggaatgtaagtaattcaagcagc 1559
Db 600 TAGGTGTGCTGGCCGGAATACAGAGTGTATCTCGAGTGACGCTACCTTGTACTG 659
Oy 1560 gattatcgacaatagag 1578
Db 660 GATTAGGCAACAACCTGG 678

RESULT 12
US-07-882-202A-3
; Sequence 3, Application US/07882202A
; Patent No. 5374617
; GENERAL INFORMATION:
; APPLICANT: Morrissey, James H.
; APPLICANT: Comp, Philip C.
; TITLE OF INVENTION: Treatment of Bleeding with Modified
; TITLE OF INVENTION: Tissue Factor in Combination with FVIIa
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Richards, Medlock & Andrews
; STREET: 1201 Elm Street, Suite 4500
; CITY: Dallas
; STATE: Texas
; COUNTRY: US
; ZIP: 75270-2197
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/882,202A
; FILING DATE: 13-MAY-1992
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Hansen, Eugenia S.
; REGISTRATION NUMBER: 31,966
; REFERENCE/DOCKET NUMBER: OMR# B34290
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 214-939-4500
; TELEFAX: 214-939-4600
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1440 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHEICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; TISSUE TYPE: Blood
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 36..1433
; OTHER INFORMATION: /note= "Coding portion of human
; OTHER INFORMATION: factor VII cDNA"
US-07-882-202A-3
Query Match 5.5%; Score 95.4; DB 1; Length 1440;

```

```

Best Local Similarity 47.9%; Pred. No. 1.4e-16;
Matches 353; Conservative 0; Mismatches 366; Indels 18; Gaps 2;
Oy 865 cgcagagagcagattgttggcgagagcgcgcctccggggccttgcccggaagtc 924
Db 660 CCCCAGGCGGATGTGGGGGCAAGGTGTCCCCAAAGGGAGGTCTCAATGAGCGGTC 719
Oy 925 agcttgacgtccagaagctccacgtgtgctgagagctccatcatcaccgcagtgatc 984
Db 720 CTGTTGTTGTAATGAGACTCAGTTGTGTGGGGGACCTCATCAACCATCTGGGGTG 779
Oy 985 gtgacagccgcccactgtgtgaaacaccttaacaatcattgcatgtgagcattt 1044
Db 780 GTCTCGCGCCGCTGCTGTTGACAAATCAAGAACTGAGGAACTGATCCGGTGTCTG 839
Oy 1045 gcgggagatttgagacaatcttcatgtctgtatggagccgagatacaagtagaanaagtg 1104
Db 840 GCGGAGCAGACCTCAGCAGCAGCAGGAGATGAGCAGAGCGCGGGTGGCGAGGTC 899
Oy 1105 attctcatccaatattatgactccaagaagaacaatgactgtgctgtatgaagtc 1164
Db 900 ATCATCCCGCAGCAGTACGTCCCGGCAACCAACAGACATGCGCTGTCCGCTG 959
Oy 1165 cagaagcctctgacttcaaggaactagtgaaacagtggtgtgtcccaa-----c 1215
Db 960 CACACCCCGGTGCTCTACTACCATGTGTGCTCCCTGCTGCTCCGAAACGAGCTTC 1019
Oy 1216 ccaggatgattgtcgaagcgaagcagctgtgatttccgggttgaggggcagcag 1275
Db 1020 TCTGAGAGGACGCTGGCTTCGTGCGCTTCATTTGTCAGCGGGTGGGCGACGTGCTG 1079
Oy 1276 gagaagaggaagcctcagaagtgctgaagcgtgcgaagtgcttcatatgagacaag 1335
Db 1080 GACCGGCGCGCCACGCGCTGAGCTCATGTGTCACAGCTCCCGGCTGATGACCCAG 1139
Oy 1336 agatgc-----aacgcgatatgtctatgacaactgatatcaacagcatgact 1386
Db 1140 GACTGCTTCACAGACTCAGCAGAGGTGGAGACTCCCAATATATCAGAGTACATGTTTC 1199
Oy 1387 tgtgcccgtctcctgcaagggagcgtcgtcttgcgaaggtgagagtcctctg 1446
Db 1200 TGTGCCGCTACTCGGATGACAGAAAGACTCTGCAAGGGGACAGTGGAGCCACAT 1259
Oy 1447 gtcaactcgaaacaatactgtgtgctgataagggatatacaagcttggtgtctgt 1506
Db 1260 GCCACCACCTACCGGGGCGACGTGTACTGACGGGCATGTCAGCTGGGGCCAGGGCTGC 1319
Oy 1507 gccaaagcttacagaccaggagtgtaacgggaatgtaagtaattcaagcagattat 1566
Db 1320 GCAACCGTGGCCACTTGTGGGTGTACACCAAGGCTCTCCCATACATCGAGTGGCNA 1379
Oy 1567 cgaacaatgagggcaga 1583
Db 1380 AAGCTCATGCGCTGAGA 1396

RESULT 13
US-08-021-615A-3
; Sequence 3, Application US/08021615A
; Patent No. 5504064
; GENERAL INFORMATION:
; APPLICANT: Morrissey, James H.
; APPLICANT: Comp, Philip C.
; TITLE OF INVENTION: Treatment of Bleeding with Modified
; TITLE OF INVENTION: Tissue Factor in Combination with an Activator of
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Richards, Medlock & Andrews
; STREET: 1201 Elm Street, Suite 4500
; CITY: Dallas
; STATE: Texas
; COUNTRY: US

```



```

; NAME/KEY: CDS
; LOCATION: 36..1433
; OTHER INFORMATION: /note="Coding portion of human
; OTHER INFORMATION: factor VII cDNA"
US-08-321-777-3

```

```

Query Match      5.5%; Score 95.4; DB 1; Length 1440;
Best Local Similarity 47.9%; Pred. No. 1.4e-16;
Matches 353; Conservative 0; Mismatches 366; Indels 18; Gaps 2;

```

```

QY 865 cgcacagagagatgtgtggcgagagcgctcccgggcccgccctggcagctc 924
DB 660 CCCCAAGGCCGGAATGTGGGGGCAAGGTGTGCCCAAGGGGAGTGTCCATGCGCAGGTC 719
QY 925 agccttcagctcagaagctcagctgtgagagctcctcctcctcctcctcctcctc 984
DB 720 CTGTTGTGTGTAATGAGACTCAGTTGTGTGGGGGAGCCCTGATCAACACATCTGGGTG 779
QY 985 gtgacagcgcgcacactgcgtggaacacactctaacatcctatgcatgtgacgcatc 1044
DB 780 GTCTCCGCGCGCCACTGTTTCGCAAAATCAAGAACTGAGAGAACTGATCGCGGTGCTG 839
QY 1045 gcgggagatttgagacaactctcctatgtctatgagccgagataccagtagaagaagtg 1104
DB 840 GCGGAGCACGACCTCAGCAGCAGCAGGAGATGAGCAGACCGCGCGGTGCGCAGGTC 899
QY 1105 attctcattcaaatatgtactcctcagaagacaaagacatgcatgctgtagaagctg 1164
DB 900 ATCATGCCAGCAGCAGTACTCCCGGGCACCACCAACGACATCGCGCTGCCCTCG 959
QY 1165 cagaagccttcgacttcaagcagcagctagtgaaacagtgctgtcgcacaa-----c 1215
DB 960 CACGAGCCCGGTGCTCTCACTACCATGTGTGTCCTCTGCTGCTGCCGACGAGAGCTTC 1019
QY 1216 cgaagcattgtctgagcagcagaagcagctctgtgtagattccgggtgggggcccacag 1275
DB 1020 TCTGAGAGGAGCGCTGCGCTTTCGCGCTTCTCATTTGTGTAGCGGGGCGGACGCTGTG 1079
QY 1276 gagaagagagaaagcctcagaagctgtcgaagcgtctgcacagctctcctcctcctc 1335
DB 1080 GACCGTGGCGCCAGCGGCGCTGAGAGTCAATGTGTGCTCAAGTGTGCTGAGTGTGAG 1139
QY 1336 agatgc-----aacagcagatatgtctatgacaacctgataccaccagcatgattc 1386
DB 1140 GACTGCTGTCACAGTCAAGGAGGTGGAGAGCTCCCAAAATATCACGAGTACATGTTTC 1199
QY 1387 tgtgcagagctctcctgagagagagctgatactctgtccagagtgtagagagcctctg 1446
DB 1200 TGTGCGGCTACTCGGATGGCAGCAAGACCTCTGCAAGGGGAGACAGTGGAGGCCACAT 1259
QY 1447 gtacattcgaagaacaatactgtgtgctgataagggatatacagaagctgggtctgtgctg 1506
DB 1260 GGCACCCACTACCGGGGCGAGTGTGATCTGACGGGCGATGTCAGTGGGGGCCAGGGCTTC 1319
QY 1507 gccaaagcttaagacagagagtgtaggggaatgtgattgttcaagcagtgattat 1566
DB 1330 GCAACCGGTGGGCGCACTTGTGGGTGTACACACAGGAGTCTCCAGTACATCGATGGCTGCAA 1379
QY 1567 cgacaatatgagggcaga 1583
DB 1380 AAGCTCATGCGCTCAGA 1396

```

```

RESULT 15
US-09-009-217-13
; Sequence 13, Application US/09009217
; Patent No. 6132729
; GENERAL INFORMATION:
; APPLICANT: Thorpe, Phillip E.
; APPLICANT: King, Steven W.
; APPLICANT: Gao, Boning
; TITLE OF INVENTION: COMBINED TISSUE FACTOR AND

```

```

; TITLE OF INVENTION: CHEMOTHERAPEUTIC METHODS AND COMPOSITIONS FOR COAGULATION
; TITLE OF INVENTION: AND TUMOR TREATMENT
; NUMBER OF SEQUENCES: 27
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold, White & Durkee
; STREET: P.O. Box 4433
; CITY: Houston
; STATE: Texas
; COUNTRY: USA
; ZIP: 77210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/009 217
; FILING DATE: Concurrently Herewith
; CLASSIFICATION:
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US 60/042,427
; FILING DATE: 27-MAR-1997
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US 60/036,205
; FILING DATE: 27-JAN-1997
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US 60/035,920
; FILING DATE: 22-JAN-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Hibler, David W.
; REGISTRATION NUMBER: 41,071
; REFERENCE/DOCKET NUMBER: UTSID:536
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 512/418-3000
; TELEFAX: 512/474-7577
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1440 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-09-009-217-13

```

```

Query Match      5.5%; Score 95.4; DB 3; Length 1440;
Best Local Similarity 47.9%; Pred. No. 1.4e-16;
Matches 353; Conservative 0; Mismatches 366; Indels 18; Gaps 2;

```

```

QY 865 cgcacagagagatgtgtggcgagagcgctcccgggcccgccctggcagctc 924
DB 660 CCCCAAGGCCGGAATGTGGGGGCAAGGTGTGCCCAAGGGGAGTGTCCATGCGCAGGTC 719
QY 925 agccttcagctcagaagctcagctgtgagagctcctcctcctcctcctcctcctcctc 984
DB 720 CTGTTGTGTGTAATGAGACTCAGTTGTGTGGGGGAGCCCTGATCAACACATCTGGGTG 779
QY 985 gtgacagcgcgcacactgcgtggaacacactcttaacaatcctatgcatgtgacgcatc 1044
DB 780 GTCTCCGCGCGCCACTGTTTCGCAAAATCAAGAACTGAGAGAACTGATCGCGGTGCTG 839
QY 1045 gcgggagatttgagacaactctcctatgtctatgtagccgagataccagtagaagaagtg 1104
DB 840 GCGGAGCACGACCTCAGCAGCAGCAGGAGATGAGCAGAGCCGCGGTGCGCAGGTC 899
QY 1105 attctcattcaaatatgtactcctcagaagacaaagacatgcatgctgtagaagctg 1164
DB 900 ATCATGCCAGCAGCAGTACTCCCGGGCACCACCAACGACATCGCGCTGCCCTCG 959
QY 1165 cagaagccttcgacttcaagcagcagctagtgaaacagtgctgtcgcacaa-----c 1215
DB 960 CACGAGCCCGGTGCTCTCACTACCATGTGTGTCCTCTGCTGCCGAGACGAGAGCTTC 1019
QY 1216 cgaagcattgtctgagcagcagaagcagctctgtgtagattccgggtgggggcccacag 1275

```

```
Db 1020 TCTGAGAGGAGCGCTGCGCTTCGTGCGCTTCATTGGTCAGCGGCTGGGCCAGCTGCTG 1079
QY 1276 gagaaggggaagaccctcagaagtgctgaacgctgcccgaagtgcttcaltgagacacag 1335
Db 1080 GACCGTGGCGCCACGGGCTCGGAGCTCATGTGCTCAACGTGCCCGCTGATGACCCAG 1139
QY 1336 agatgc-----aacagcagatatgctctatgacaaacctgatacacaccagccatgac 1386
Db 1140 GACTGCTGACGACGACGACGAGGAGGAGTGGAGACTCCCCAAATATACGAGAGTACATGTTT 1199
QY 1387 tctgcccgtcttcctcaggggaaacgtcgaatctctgccaagggtgacagtgaggggcctctg 1446
Db 1200 TGTGCGGCTACTCGGATGGCAGCAAGGACTCCTGCAAGGGGGACAGTGGAGGCCACAT 1259
QY 1447 gtcaactcgaagaacataatctggtgctgatataggggataacaagctggggttctgctgt 1506
Db 1260 GCCACCCACTACGCGGGGACAGTGTGTACCTGTACGGGCATCGTCAGCTGGGGCCAGGGCTGC 1319
QY 1507 gccaaagcttacagaccagagtgctacggggaatgtaagtgtatcagcgaatgatat 1566
Db 1320 GCAACCGTGGGCCACTTGTGGGGTGTACACCAAGGCTCTCCAGTACATGAGTGGCTCAA 1379
QY 1567 cgacaaatgagggcaga 1583
Db 1380 AAGCTCATGCGCTCAGA 1396
```

Search completed: September 26, 2001, 09:22:54
Job time: 81 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 26, 2001, 09:21:05 : Search time 34.76 Seconds

(without alignments)
858.084 Million cell updates/sec

Title: US-09-615-285-2

Perfect score: 2717

Sequence: 1 MALNSGSPPAIGPYENHGY.....YGVNMFPTDIYROMRADG 492

Scoring table:

BLOSUM62

Gapop 10.0 , Gapept 0.5

Searched: 412676 seqs, 60623988 residues

Total number of hits satisfying chosen parameters: 412676

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database :

A.Geneseq_0601:*
1: /SID58/gcgdata/geneseq/geneseq/AA1980.DAT:*
2: /SID58/gcgdata/geneseq/geneseq/AA1981.DAT:*
3: /SID58/gcgdata/geneseq/geneseq/AA1982.DAT:*
4: /SID58/gcgdata/geneseq/geneseq/AA1983.DAT:*
5: /SID58/gcgdata/geneseq/geneseq/AA1984.DAT:*
6: /SID58/gcgdata/geneseq/geneseq/AA1985.DAT:*
7: /SID58/gcgdata/geneseq/geneseq/AA1986.DAT:*
8: /SID58/gcgdata/geneseq/geneseq/AA1987.DAT:*
9: /SID58/gcgdata/geneseq/geneseq/AA1988.DAT:*
10: /SID58/gcgdata/geneseq/geneseq/AA1989.DAT:*
11: /SID58/gcgdata/geneseq/geneseq/AA1990.DAT:*
12: /SID58/gcgdata/geneseq/geneseq/AA1991.DAT:*
13: /SID58/gcgdata/geneseq/geneseq/AA1992.DAT:*
14: /SID58/gcgdata/geneseq/geneseq/AA1993.DAT:*
15: /SID58/gcgdata/geneseq/geneseq/AA1994.DAT:*
16: /SID58/gcgdata/geneseq/geneseq/AA1995.DAT:*
17: /SID58/gcgdata/geneseq/geneseq/AA1996.DAT:*
18: /SID58/gcgdata/geneseq/geneseq/AA1997.DAT:*
19: /SID58/gcgdata/geneseq/geneseq/AA1998.DAT:*
20: /SID58/gcgdata/geneseq/geneseq/AA1999.DAT:*
21: /SID58/gcgdata/geneseq/geneseq/AA2000.DAT:*
22: /SID58/gcgdata/geneseq/geneseq/AA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2717	100.0	492	21	AAV77726
2	2717	100.0	492	21	AAV44406
3	2709	99.7	492	21	AAV92050
4	2696	99.2	492	21	AAV36901
5	2696	99.2	492	21	AAV57280
6	1540	56.7	283	21	AAV81492
7	888	32.7	454	21	AAV32246
8	885	32.6	452	20	AAV41694
9	884	32.5	453	21	AAV44250
10	742.5	27.3	327	21	AAV72093
11	676.5	24.9	435	20	AAV06437

12	676.5	24.9	435	22	AAV72558	Human seripancrin
13	674	24.8	432	21	AAV99417	Human PRO1570 (UNO
14	674	24.8	432	22	AAV87581	Human PRO1570. HO
15	674	24.8	432	22	AAV66166	Protein of the Inv
16	671.5	24.7	432	22	AAV72559	Human seripancrin
17	666.5	24.5	480	21	AAV08950	Human secreted pro
18	665.5	24.5	445	21	AAV11698	Mouse serine prote
19	660	24.3	798	15	AAV57283	Bovine enterokinase
20	651.5	24.0	457	21	AAV11699	Human serine prote
21	649.5	23.9	248	21	AAV43572	Human cancer assoc
22	648.5	23.9	414	21	AAV08912	Human secreted pro
23	601	22.1	1042	21	AAV44426	Human serine prote
24	588	21.6	311	21	AAV11697	Mouse serine prote
25	587.5	21.6	273	21	AAV11696	Mouse serine prote
26	577.5	20.6	1113	21	AAV44427	Mouse Serine prote
27	558.5	20.6	683	21	AAV19551	Human matrilysin
28	558.5	20.6	762	21	AAV90284	Human peptidase, H
29	558.5	20.6	855	20	AAV06671	Tumour antigen der
30	558.5	20.6	855	21	AAV19552	Human membrane-ty
31	558.5	20.6	855	22	AAV35465	Human membrane-ty
32	558	20.5	296	21	AAV72108	Human serine prote
33	558	20.5	372	21	AAV72092	Human serine prote
34	544	20.0	238	21	AAV11695	Mouse, serine prote
35	539.5	19.9	416	20	AAV43325	Mouse hepsin prote
36	539.5	19.9	416	20	AAV96812	A mouse serine pro
37	538.5	19.8	802	20	AAV41710	Human PRO618 (UNO3
38	538.5	19.8	802	21	AAV44266	Human PRO618 (UNO3
39	538.5	19.8	802	21	AAV24052	Human PRO618 (UNO3
40	534	19.7	356	19	AAV46917	Human PRO618 (UNO3
41	527.5	19.4	418	17	AAV89435	Amino acid sequenc
42	527.5	19.4	418	20	AAV29498	Trypsin-like enzym
43	527.5	19.4	418	20	AAV29501	Human lung tumour
44	527.5	19.4	418	20	AAV29502	Human lung tumour
45	527.5	19.4	418	21	AAV44437	Human lung tumour

ALIGNMENTS

RESULT 1
AAV77726
ID AAV77726 standard; Protein: 492 AA.
XX
AC AAV77726:
XX
DT 12-MAY-2000 (first entry)
XX
DE Human tumour suppressor TMPRSS2 polypeptide.
XX
KW Tumour suppressor gene; TMPRSS2; cancer; human; drug design;
KW gene therapy; protein therapy.
XX
OS Homo sapiens.
XX
PN MO20000605-A1.
XX
PD 06-JAN-2000.
XX
PE 29-JUN-1999; 99MO-US14622.
XX
PR 29-JUN-1998; 98US-0091044.
XX
PA (MYRI-) MYRIAD GENETICS INC.
XX
PI Wong AKC, Tavtigian SV, Teng DHF;
XX
DR WPI: 2000-170914/15.
XX
DR N-PSDB: AAV87786.
XX
PT Novel tumor suppressor TMPRSS2 used for the diagnosis and prognosis of
XX human cancer -
XX
PS Claim 55; Page 77-79; 89pp; English.

XX The invention provides a new tumour suppressor gene, designated TMRSS2.
 CC The TMRSS2 polynucleotides and polypeptides can be used in methods for
 CC diagnosing and prognosing predisposition to cancer in humans. The
 CC polypeptides may also be used in assays to screen for compounds with
 CC anti-cancer or therapeutic properties. The polypeptides are also useful
 CC for rational drug design. The TMRSS2 polynucleotides and polypeptides
 CC may be used for gene therapy and protein therapy. The present sequence
 CC represents the TMRSS2 polypeptide.

XX Sequence 492 AA:

Query Match 100.0%; Score 2717; DB 21; Length 492;
 Best Local Similarity 100.0%; Pred. No. 2.9e-203;
 Matches 492; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MALNGSPPAIGPYENHGYOPENPYPAQPTVYVYEVHNAQYRSPVQYAPRVLTQA 60
 DB 1 malngsppaigpyenhyqpenypaqptvvyevhnpqyypvqyaprvltqa 60
 QY SNPVCTOPKSPSGVCTSKTKKALCITLTGTFVGAALAGLLKFMGSCNSGIEC 120
 DB 61 snpvcctopkspsgvctsktkkalcitltgtflvgaalaagllkfmgskcnsגיע 120
 QY 121 DSSGTCINPNCMDGVSHCPGEGEDENRCVRLYGNPFILOVYSSQRKSWHPVCQDDMNEY 180
 DB 121 dssgtcinpncmdgvshcpgeddenrcvrllygnpfllqvysqrkswhpvcqddwnehy 180
 QY 181 GGAACRDMGYKKNFYSSQGITVDGSGTSMKLNNTSAGNVDIKKLYHSDACSKAVVSLR 240
 DB 181 ggaacrdmgykknfyssqgitvdsgstsmklnntsaagnvdlykklyhsdacsckavvslr 240
 QY 241 CIACGVNLNLSRQSRIVGESALPGAMPQVSLHVNHVCGSITTPMIVTAHCVEK 300
 DB 241 ciacgvnlinsrqsriygesalpgampqvslhvnhvvcgsittpmivtaahcvek 300
 QY 301 PLNNPWHMTAFAGILRQSFMEYAGYQVEKVISHPNYDSKTKNNDIALMKLOKPLTFNDL 360
 DB 301 plnnpwhmtafagilrqsfmeyagyeqvshpnydsktkndialmklqkpltfndl 360
 QY 361 VAPVCLPNPGMLOPEQLCWMISGMATEEKGTSFVLNAKVLLETQRCSRYVYDNL 420
 DB 361 vapvclpnpgmlopeqlcwmisgmateekgtsfvlnaakvlltqrcsryvyydnl 420
 QY 421 TPAMICAGFLQGNVDSGCGDGGPVTYSKNNITMWLIGDTSMGSCAKAYRPGYGNVMVF 480
 DB 421 tpamicagflqgnvdsccgdsgrplvtysknnlwmllgdtswsgcakayrpgygnvmvf 480
 QY 481 TDWITRQHMADG 492
 DB 481 tdwityrqrmdg 492

RESULT 2

AAAY44406
 ID AAY44406 standard; Protein: 492 AA.

AC AAY44406;

DT 22-MAR-2000 (first entry)

DE Human 20P1F12-GTCC protein.

XX 20P1F12; TMRSS2; androgen; serine protease; 20P1F12-GTCC; cancer;

KW transmembrane protein; colon; prostate; prostate tumour;

OS Homo sapiens.

XX WO962942-A2.

PD 09-DEC-1999.

XX

PF 01-JUN-1999; 99WO-US12253.

XX 01-JUN-1998; 98US-0087598.

PR 29-JUN-1998; 98US-0091474.

PR 14-APR-1999; 99US-0129521.

XX (UROC-) UROGENESIS INC.

PA (AFAR/) AFAR D E.

PA (HUBB/) HUBERT R S.

PA (LEON/) LEONG K.

PA (RAIT/) RAITANO A B.

PA (SAFE/) SAFRAN D C.

PI Afar DE, Hubert RS, Leong K, Raitano AB, Safran DC;

DR WPI: 2000-116363/10.

DR N-PSDB: AAZ29636.

XX Novel cell surface antigen useful to treat colon and prostate cancer -

PS Claim 1; Fig 1; 58pp; English.

CC The present sequence is the 20P1F12 protein (also known as the TMRSS2

CC protein) which is a prostate-specific, androgen-regulated, cell surface

CC serine protease. It is a glycosylated type II transmembrane protein with

CC an extracellular C-terminal serine protease domain, a scavenger receptor

CC cysteine-rich domain, an LDL receptor class A domain and a predicted

CC transmembrane domain. Host cells can be transformed to produce this

CC 20P1F12-GTCC, as deposited with ATCC accession number 207097).

CC Anti-20P1F12/TMRSS2 antibodies may be used as therapeutic agent for

CC prostate and colon cancers, to image prostate cancer cells and

CC prostate tumours, to identify ligands and cellular constituents that

CC bind to a 20P1F12/TMRSS2 gene product and for use as cancer vaccines.

XX Sequence 492 AA:

Query Match 100.0%; Score 2717; DB 21; Length 492;
 Best Local Similarity 100.0%; Pred. No. 2.9e-203;
 Matches 492; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MALNGSPPAIGPYENHGYOPENPYPAQPTVYVYEVHNAQYRSPVQYAPRVLTQA 60
 DB 1 malngsppaigpyenhyqpenypaqptvvyevhnpqyypvqyaprvltqa 60
 QY 61 SNPVCTOPKSPSGVCTSKTKKALCITLTGTFVGAALAGLLKFMGSCNSGIEC 120
 DB 61 snpvcctopkspsgvctsktkkalcitltgtflvgaalaagllkfmgskcnsגיע 120
 QY 121 DSSGTCINPNCMDGVSHCPGEGEDENRCVRLYGNPFILOVYSSQRKSWHPVCQDDMNEY 180
 DB 121 dssgtcinpncmdgvshcpgeddenrcvrllygnpfllqvysqrkswhpvcqddwnehy 180
 QY 181 GGAACRDMGYKKNFYSSQGITVDGSGTSMKLNNTSAGNVDIKKLYHSDACSKAVVSLR 240
 DB 181 ggaacrdmgykknfyssqgitvdsgstsmklnntsaagnvdlykklyhsdacsckavvslr 240
 QY 241 CIACGVNLNLSRQSRIVGESALPGAMPQVSLHVNHVCGSITTPMIVTAHCVEK 300
 DB 241 ciacgvnlinsrqsriygesalpgampqvslhvnhvvcgsittpmivtaahcvek 300
 QY 301 PLNNPWHMTAFAGILRQSFMEYAGYQVEKVISHPNYDSKTKNNDIALMKLOKPLTFNDL 360
 DB 301 plnnpwhmtafagilrqsfmeyagyeqvshpnydsktkndialmklqkpltfndl 360
 QY 361 VAPVCLPNPGMLOPEQLCWMISGMATEEKGTSFVLNAKVLLETQRCSRYVYDNL 420
 DB 361 vapvclpnpgmlopeqlcwmisgmateekgtsfvlnaakvlltqrcsryvyydnl 420
 QY 421 TPAMICAGFLQGNVDSGCGDGGPVTYSKNNITMWLIGDTSMGSCAKAYRPGYGNVMVF 480
 DB 421 tpamicagflqgnvdsccgdsgrplvtysknnlwmllgdtswsgcakayrpgygnvmvf 480

OY 481 TDWYROMRADG 492
|||||
Db 481 tdwlyrqmrads 492

RESULT 3

AAV92050
ID AAV92050 standard; Protein; 492 AA.
XX
AC AAV92050;

DT 01-AUG-2000 (first entry)

DE HRPc6/7 polypeptide from androgen-inducible gene clone.

KM Androgen inducible; testosterone; prostate cancer; cytostatic;
KW TMPRSS2; diagnosis.

OS Homo sapiens.

PN W0200018961-A2.

PD 06-APR-2000.

PE 30-SEP-1999; 99W0-US22535.

PR 30-SEP-1998; 98US-0163759.
PR 30-SEP-1998; 98US-0164159.

PA (MILL-) MILLENNIUM PHARM INC.

PI Macbeth KJ, Shyjan AW;

DR WPI: 2000-293182/25.
DR N-PSDB: AAA08803.

PT Novel methods for identifying compounds for treating prostate cancer
PT comprising measuring the level of expression or activity of 1 or more
PT of 11 genes or their products

XX Claim 2; Fig 3; 108pp; English.

XX This protein is encoded by a gene which is androgen (e.g. testosterone)
CC inducible in androgen-dependent prostate cancer cells (e.g. LNCaP cells)
CC and constitutively expressed in androgen-independent prostate cancer
CC cells (e.g. LNCaP cells). Agents which decrease the expression or
CC activity of these clones may slow or arrest the growth of prostate cancer
CC cells or may kill them. HRPc6/7 can be obtained from the sequence of
CC the known gene for TMPRSS2. A compound useful for treating prostate
CC cancer can be identified in a novel method comprising measuring the
CC expression level or activity, of HRPc6/7, 6/7, 8, 9, 10, 13, 14, 15,
CC 19, or peripheral-type demodilazepine receptor (PBR) in a cell, in the
CC presence and absence of a test compound. The sequences may also be used
CC in diagnosis of prostate cancer and to determine efficacy of treatment
CC for prostate cancer.

XX Sequence 492 AA:

Query Match 99.7%; Score 2709; DB 21; Length 492;
Best Local Similarity 99.8%; Pred. No. 1.2e-202;
Matches 491; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 MALNCGSPAIPIPIYENHGYOENPPAOPTYPPYEVHPAOTYSPVQVAPRLTQA 60
|||||

Db 1 malnsgspaiipiyenhyqpenypaqvtvpyevhpqyypvqyaprvltqa 60
|||||

OY 61 SNPVVCTOKSPSGVCTCTKTKALCTLTGLGFTVGAALAGLTKFMGSKCSNSGIFC 120
|||||

Db 61 snpvcvtokspsgvctctkktkalcitltglftvgaalaglltkfmgskcsnsjgle 120
|||||

OY 121 DSSGTCINPSNMCDGVSHCPGGEDENRVRLYGNPFILOVSSORSKSMHPVCODDMNENY 180
|||||

Db 121 dssgtcinspncdgvshcpbggedenrcvrllygsnfllyvssqgrkswmpvcoddmneny 180
|||||

OY 181 GRAACRDMGCKNNFYSSQGIYDDSGSTSPFKLNTSAGNNDIYKLLHSDACSSKAYVSLR 240
|||||

Db 181 graacrdmgyknnfyssqgiyddsgstsfmk lntsaqnvdlykllhsdacsakavvs1r 240
|||||

OY 241 CIACGVNLNSSRQSRIVGSESLPGAMPVOSLHVONVHVCAGSIITPEWIVAAACVEK 300
|||||

Db 241 ciacgvnlssrsqsrivvgesalpgawpovslhvgnvhvcagsiltpewivaaacvek 300
|||||

OY 301 PLNNPWHMTAFACILROSEMFYAGYQVEKVISHPNVDSKTKNNDIALMLKQPLTFNDL 360
|||||

Db 301 plnpwhmtafagilrsgfmfyagyyvekvlshpnvdsstkndialmlkqltfnl 360
|||||

OY 361 VKPVCIPNPGMMLQPOLCHISGMCATEEKGKTSEVLNAAKVLLITQRCNSRYVDNLI 420
|||||

Db 361 vkpvc1pnpgmmlqpeqlcwisgwaeteekgtsevlnaakvllitqrcnsryvndli 420
|||||

OY 421 TPAMICAGFLOGNVDSGCGSGPLVTSKNNIMWLLIGDTSWGSCKAKAYRPGYGVNMYF 480
|||||

Db 421 tpamicagflqgnvdsccgsgplvtsknnimwlligdtswsgcakayrpgygvnmvf 480
|||||

OY 481 TDWYROMRADG 492
|||||

Db 481 tdwlyrqmrads 492

RESULT 4

AAAB36901
ID AAB36901 standard; Protein; 492 AA.
XX

AC AAB36901;

DT 26-FEB-2001 (first entry)

DE Human TMPRSS2 protein.

KW Prostate specific androgen regulated protein; ARSDRL; TMPRSS2;
KW PART-1; neoplastic.

OS Homo sapiens.

PN W0200065067-A2.

PD 02-NOV-2000.

PE 21-APR-2000; 2000W0-US10920.

PR 23-APR-1999; 99US-0130778.
PR 30-AUG-1999; 99US-0151585.
PR 30-DEC-1999; 99US-0174003.
PR 24-JAN-2000; 2000US-0177751.

PA (UNITV) UNITV WASHINGTON.

PI Nelson PS, Hood L, Lin B;

DR WPI: 2000-679676/66.
DR N-PSDB: AAC83325.

PT Polynucleotide encoding prostate specific androgen regulated
PT polypeptides and inhibitor of the peptides useful for treating or
PT reducing the progression of prostate neoplastic condition in an
PT individual -
XX Claim 63; Page 86-88; 121pp; English.

XX The present invention relates to prostate specific androgen regulated
XX proteins. The invention may be used to determine an expression level
XX of the prostate-specific proteins ARSDRL, TMPRSS2, or PART-1 in a
XX fluid sample or prostate cell sample from an individual. It may also
XX be used for diagnosing and predicting the susceptibility of a

CC prostate neoplastic condition in an individual. Inhibitors of the
CC proteins are useful for treating or preventing the progression of a
CC prostate neoplastic condition.

Sequence 492 AA:

Query Match 99.2%; Score 2696; DB 21; Length 492;
Best Local Similarity 98.8%; Pred. No. 1.2e-201;
Matches 486; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

OY 1 MALNSGSPPAICPYENHGYOPENPYPAQIVPVYVEYHPAQYRPSVPQYAPRVLTA 60
DB 1 malnsghppaipyenhyqpenpypaqivpvtyevhpqyypvpqyaprvltga 60
OY 61 SNPVCTOPKSPSGVTCSTKTKKALCITLTGTFVLGALAGLMLKFMGSKCSNSGIC 120
DB 61 snpvctopkspsgvctcstkktkalcitltgtflvgaalaagllwktmgskcsnsgiec 120
OY 121 DSSGTCINPSNMCDVSHCPGGEDENRVRLYGPNEFLQVYSSQRKSMHPVCODDMNENY 180
DB 121 dssgtcinsnmcdvshcpggedenrcvrllygpnflqmysqkswhpvcoddmneny 180
OY 181 GRAACRDMCKYKNNFYSSGCIYVDDSGTSPMKLNTSAGNVDIKKILYHSDACSKAVVSLR 240
DB 181 graacrdmgyknnfyssgciyvdsgstfmklnsaagnvdiykklyhsdacskskavslr 240
OY 241 CIACGVNLNSSRQSRIVGESALPGAMPQVSLHYQNVHVCGSITTPMWITAAHCVEK 300
DB 241 ciacgvnlssrqsriygesalpgampqvslyhqnvhvcgsiltpewvtaahcvek 300
OY 301 PLNNPWHMTAFAGILRQSFMEYAGYQVEKVISHPNYDSKTKNNDIALMKLQKPLTFNDL 360
DB 301 plnpwhmtafagilrqsfmfyagvyqkvishpnydsktkndialmkkqlptlfnl 360
OY 361 VKPVCPLPNPMMLOPEQLCWMISGKATEEKGKTSYLNAKVLLETORCNSRYVDNL 420
DB 361 vkpvcplpnpmmlpbeqlcwmisgwateekgtsevlaakvllletqrcnsryydnll 420
OY 421 TPAMICAGFLQGNVDSGCGDGGPLVTSKNNIMWMLIGDTSMSGCAKAYRPVYGNVMVF 480
DB 421 tpamicagflqgnvdscgdsqgplvtstnnlmwlllgdtsmsgcakayrpvygnvmvf 480
OY 481 TDMYIROMRADG 492
DB 481 tdwlyrqnkang 492

RESULT 5

AA57280
ID AA57280 standard; Protein: 492 AA.

AC AA57280;

XX 06-JUN-2000 (first entry)

DE Ovr115 homolog protein.

XX CSC; cancer specific gene; cancer; gynecologic cancer; ovarian; breast;

KW endometrial; uterine; Lung; cytotoxic.

OS Homo sapiens.

XX MO200012758-A1.

PD 09-MAR-2000.

XX 01-SEP-1999; 99WO-US19655.

XX 02-SEP-1998; 98US-0098880.

XX (DIAD-) DIADEXUS LLC.

PI Salceda S, Sun Y, Recipon H, Caffrey R;
XX WPI: 2000-256657/22.
DR N-PSDB: AA290478.

PT Diagnosing, staging, monitoring, imaging and treating cancer especially
PT gynecological cancers e.g. breast, ovarian cancer and lung cancer,
PT involves measuring cancer specific gene levels in cells and body fluids

PS Disclosure: Page 52-54; 58pp; English.

XX The invention relates to detecting, diagnosing metastasis and staging
CC cancer by measuring levels of cancer specific genes (CSG) in cells,
CC tissues or body fluids. Their remission and progression, decreases and
CC increases in CSG levels, is also monitored, by periodic sample analysis.
CC The methods are useful for detecting cancers, especially gynecologic
CC cancers which include ovarian, breast, endometrial and uterine cancer
CC and lung cancer. Antibodies against the CSGs labeled with paramagnetic
CC ions or a radioactive agent is useful for imaging cancer and when conjugated
CC with a cytotoxic agent are useful for treating cancer. The present
CC sequence represents a Ovr115 homolog protein, that can be used
CC for the detection of the various cancers.

Sequence 492 AA:

Query Match 99.2%; Score 2696; DB 21; Length 492;
Best Local Similarity 98.8%; Pred. No. 1.2e-201;
Matches 486; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

OY 1 MALNSGSPPAICPYENHGYOPENPYPAQIVPVYVEYHPAQYRPSVPQYAPRVLTA 60
DB 1 malnsghppaipyenhyqpenpypaqivpvtyevhpqyypvpqyaprvltga 60
OY 61 SNPVCTOPKSPSGVTCSTKTKKALCITLTGTFVLGALAGLMLKFMGSKCSNSGIC 120
DB 61 snpvctopkspsgvctcstkktkalcitltgtflvgaalaagllwktmgskcsnsgiec 120
OY 121 DSSGTCINPSNMCDVSHCPGGEDENRVRLYGPNEFLQVYSSQRKSMHPVCODDMNENY 180
DB 121 dssgtcinsnmcdvshcpggedenrcvrllygpnflqmysqkswhpvcoddmneny 180
OY 181 GRAACRDMCKYKNNFYSSGCIYVDDSGTSPMKLNTSAGNVDIKKILYHSDACSKAVVSLR 240
DB 181 graacrdmgyknnfyssgciyvdsgstfmklnsaagnvdiykklyhsdacskskavslr 240
OY 241 CIACGVNLNSSRQSRIVGESALPGAMPQVSLHYQNVHVCGSITTPMWITAAHCVEK 300
DB 241 ciacgvnlssrqsriygesalpgampqvslyhqnvhvcgsiltpewvtaahcvek 300
OY 301 PLNNPWHMTAFAGILRQSFMEYAGYQVEKVISHPNYDSKTKNNDIALMKLQKPLTFNDL 360
DB 301 plnpwhmtafagilrqsfmfyagvyqkvishpnydsktkndialmkkqlptlfnl 360
OY 361 VKPVCPLPNPMMLOPEQLCWMISGKATEEKGKTSYLNAKVLLETORCNSRYVDNL 420
DB 361 vkpvcplpnpmmlpbeqlcwmisgwateekgtsevlaakvllletqrcnsryydnll 420
OY 421 TPAMICAGFLQGNVDSGCGDGGPLVTSKNNIMWMLIGDTSMSGCAKAYRPVYGNVMVF 480
DB 421 tpamicagflqgnvdscgdsqgplvtstnnlmwlllgdtsmsgcakayrpvygnvmvf 480
OY 481 TDMYIROMRADG 492
DB 481 tdwlyrqnkang 492

RESULT 6

AA581492
ID AA581492 standard; Protein: 283 AA.

AC AA581492;

```

XX 18-JUL-2000 (first entry)
DT
XX Human prostate-associated protease (HUPAP).
DE
XX Human prostate-associated protease; HUPAP; kallikrein; serine protease;
KW gastrointestinal disorder; cancer; prostate disorder.
XX
OS Homo sapiens.
FH Key Location/Qualifiers
FT Modified-site 4 /note= "N-glycosylated"
FT Misc-difference 235 /label= unknown
FT /note= "Encoded by CMT"
FT
XX US6043033-A.
XX
XX 28-MAR-2000.
XX
XX 27-FEB-1997; 97US-0807151.
XX
XX 27-FEB-1997; 97US-0807151.
XX
XX (INCY-) INCYTE PHARM INC.
XX
XX Bandman O, Lal P;
XX
XX WPI; 2000-282523/24.
XX
XX N-PSDB; AAA12975.
XX
XX Polynucleotide encoding human prostate-associated protease useful for
PT diagnosing and treating cancers; prostate disorders and
PT gastrointestinal disorders
XX
XX Claim 1; Fig 1A-B; 27pp; English.
XX
XX This sequence represents human prostate-associated protease (HUPAP). cDNA
XX encoding HUPAP was initially identified in a spinal cord cDNA library,
XX the cDNA encoding this sequence representing a consensus of overlapping
XX and/or extended nucleic acid sequences from spinal cord, prostate tumour
XX and colon cDNA libraries. HUPAP is a serine protease with structural and
XX functional homology with bovine enterokinase, human pancreatic kallikrein
XX and African rat renal kallikrein, sharing 38% homology with bovine
XX enterokinase. In addition, HUPAP is related to prostate-specific antigen
XX (PSA), a kallikrein which is a highly sensitive marker for prostate
XX cancer. HUPAP and nucleic acids encoding it are useful for the diagnosis,
XX prevention and treatment of gastrointestinal disorders such as ulcerative
XX colitis, pancreatitis, cancers, and prostatic disorders. HUPAP agonists
XX may be used to treat gastrointestinal disorders, and HUPAP antagonists
XX and inhibitors may be used to suppress excessive cell proliferation,
XX which is of use in cancer therapy. HUPAP nucleic acids are also useful
XX for generating hybridisation probes which may be used for mapping
XX naturally occurring genomic sequences.
XX
XX Sequence 283 AA:
SO

```

```

Db 121 kvishpnydstkndialmklqkpltfndlvkpvclpnpmmilqpeqlcwisgagtee 180
Qy 390 KGRSEVLNAAKVLITFQRNSRYVDNLTPMICAQFLQGVNDSCGSGSPLYTSK 449
Db 181 kgktsevlinaakvllletqrnsryvndltlpamiaagflqgvndscgsgsylvtsk 240
Qy 450 NNIWMILIGDTSWGSCKAKAYRPGYGNVWPTDVIYRQMRADG 492
Db 241 nniwmiligtswsgcakayrpgygnvwptdviyrmradg 283

RESULT 7
ID AAB32246 standard; Protein; 454 AA.
XX
XX AAB32246;
XX
XX 11-JAN-2001 (first entry)
XX
XX Tumour associated differentially-expressed gene 12 protein sequence.
DE
XX
XX Transmembrane serine protease; TADG-12; chromosome 17; vaccination;
KW tumour associated differentially-expressed gene 12; cytostatic; human;
XX malignant hyperplasia; cancer; ovary; breast; lung; colon; prostate.
XX
XX Homo sapiens.
XX
XX WO200052044-A1.
XX
XX 08-SEP-2000.
XX
XX 02-MAR-2000; 2000WO-US05612.
XX
XX 03-MAR-1999; 99US-0261416.
XX
XX (UYAR-) UNIV ARKANSAS.
XX
XX O'Brien TJ, Underwood LJ;
XX
XX WPI; 2000-533263/48.
XX
XX N-PSDB; AAA93842.
XX
XX DNA fragment encoding tumor associated differentially-expressed gene 12
PT protein used for diagnosing and treating malignant hyperplasia and
PT cancers including ovarian cancer
XX
XX Claim 3; Figure 4; 11pp; English.
XX
XX This invention relates to a novel transmembrane serine protease called
XX tumour associated differentially-expressed gene 12 (TADG-12). TADG-12 is
XX located on chromosome 17. Sequences AAA93842-A93845 and AAB32246-B32249
XX represent human TADG-12 cDNA and their corresponding protein sequences.
XX A splice variant of TADG-12 (TADG-12V) leads to a truncated protein
XX product. TADG-12 is overexpressed in ovarian carcinomas. TADG-12
XX exhibits cytoskeletal activity, and can be used in vaccines and in gene
XX therapy. TADG-12 nucleotide and protein sequences are used in the
XX diagnosis of malignant hyperplasia and cancers of the ovary, breast,
XX lung, colon, prostate and other cancers where TADG-12 is overexpressed.
XX TADG-12 is particularly used as tumour marker for early disease
XX diagnosis. TADG12 proteins or fragments can be used to vaccinate an
XX individual with cancer, suspected of having a cancer or at risk of
XX getting cancer. Sequences AAA93846-A93853 represent PCR primers used for
XX amplifying the TADG-12 cDNA sequence, and in the quantitative analysis of
XX TADG-12 mRNA. AAB32250 represents a peptide fragment of TADG-12, used to
XX create anti-TADG-12 antibodies. Sequences AAB32251-B32269 represent
XX TADG-12 peptides which target HLA, and may be used in a vaccine or for
XX immune stimulation.
XX
XX Sequence 454 AA:
SO

```

Query Match 32.7%; Score 888; DB 21; Length 454;
Best Local Similarity 46.3%; Pred. No. 3.8e-61;

FT Modified-site 144
 FT /note= "protein kinase C phosphorylation site"
 FT Modified-site 148
 FT /note= "protein kinase C phosphorylation site"
 FT Modified-site 176
 FT /note= "N-glycosylated"
 FT Modified-site 197
 FT /note= "protein kinase C phosphorylation site"
 FT Modified-site 200
 FT /note= "protein kinase C phosphorylation site"
 FT Modified-site 231
 FT /note= "casein kinase II phosphorylation site"
 FT Active-site 243
 FT /note= "characteristic of serine protease"
 FT Modified-site 249
 FT /note= "cAMP- and cGMP-dependent protein kinase phosphorylation site"
 FT Modified-site 260
 FT /note= "protein kinase C phosphorylation site"
 FT Modified-site 303
 FT /note= "protein kinase C phosphorylation site"
 FT Modified-site 351
 FT /note= "protein kinase C phosphorylation site"
 FT Modified-site 360
 FT /note= "tyrosine kinase phosphorylation site"
 FT Modified-site 365
 FT /note= "protein kinase C phosphorylation site"
 FT Active-site 385
 FT /note= "characteristic of serine protease"
 FT
 FT W09936550-A2.
 FT
 FT 22-JUL-1999.
 FT
 FT 12-JAN-1999; 99WO-US00655.
 FT
 FT 16-JAN-1998; 98US-0008271.
 FT
 FT (INCY-) INCYTE PHARM INC.
 FT
 FT Bandman O, Corley NC, Guegler KJ, Hillman JL, Shah P;
 FT Tang YJ, Yue H;
 FT WPI: 1999-430616/36.
 FT N-PSDB; AAX87154.
 FT
 FT Novel human protease molecules useful in the treatment of
 FT developmental disorders and/or cancers
 FT
 FT Claim 1; Page 74-75; 90pp; English.
 FT
 FT The present sequence represents novel human protease HUPM-6, as
 FT deduced from the consensus sequence (see AAX87154) of overlapping
 FT cDNA clones obtained from various libraries. Northern analysis
 FT shows expression of HUPM-6 in gastrointestinal, and male and
 FT female reproductive cDNA libraries. Approximately 65% of these
 FT libraries are associated with neoplastic disorders, and 22% with
 FT the immune response. The invention provides 12 new human
 FT proteases, i.e. HUPM-1 to -12 (see AAX06432-43), and the
 FT polynucleotides encoding them (see AAX87149-60). Also provided are
 FT vectors, host cells and methods for producing HUPM polypeptides,
 FT as well as agonists and antagonists of HUPM. Methods for treating
 FT or preventing cell proliferative disorders and immune disorders
 FT using HUPM or HUPM antagonists are claimed.
 FT
 FT Sequence 435 AA:

Query Match 24.9%; Score 676.5; DB 20; Length 435;
 Best Local Similarity 39.1%; Pred. No. 9.8e-45;
 Matches 150; Conservative 57; Mismatches 126; Indels 49; Gaps 13;
 QY 133 CDGVSHCPGDEENRCVRLY--GP-----NFILOYSSQSKSMHPVCDQNMENYNG 181

Db 75 cdgedlcpjgedeehcvksfpegpavavrlskdrstlqyldsatgwfacafnftaala 134
 QY 182 RAACRDMGY--KNNFYSQ-----GIYDDSGSTSFMKLNTSAGVNDYIKRIYHDAQS 232
 Db 135 elacrcmgyskprtfaveipgpdldvveltenstgelmrnssg-----pcl 182
 QY 233 SKAVSLRGIACGVNLTNSRNRIGVESALPGAMPQVYSLHVQNVGCGSIITPEWIV 292
 Db 183 sgsivslhlaagel---kprvvgaeasvdswpwqslgdkqhvcgslllqphwv1 239
 QY 293 TAAHCEKPLNPMWHTAPAGILR--OSFMEYGAGYOVERVI---SHPNYDSKTKNDIAL 348
 Db 240 taahcfrkh-tdfwkvragssklgsf---pslavakillilefpmv---pkndial 291
 QY 349 MKIQKPLTFNDLVKPYCLPMPGMLOPEQLCWTSGGATFEK-GKISEVLMNAKVLIT 407
 Db 292 mkqlfpltfstgvtvripclpfdeelpatpjlwlgwftkqngykmddlqasvqids 351
 QY 408 ORCNSRYVDNLITTPAMICAGFLQGNVDSQGDGSGPLVTSKNNIMWLIGDTSWGSGCAK 467
 Db 352 trcnaddayqgevtckmmcaaglpvgvdtcggdsgplmygsdq-whvvglysvgy9cg9 410
 QY 468 AYRPGVYGNVWVETDMYRQMRAD 491
 Db 411 pstpgvytkvsaylwnlynwkae 434
 RESULT 12
 AAY72558
 ID AAY72558 standard; Protein; 435 AA.
 XX
 AC AAY72558;
 XX
 DT 02-MAY-2001 (first entry)
 XX
 DE Human seripancrin protein.
 XX
 KW Human; seripancrin; serine protease; chromosome 11q22-q23; therapy;
 KW arthritis; chronic obstructive pulmonary disorder; COPD; cancer;
 KW osteoporosis; aberrant wound healing; angiogenesis; diabetes;
 KW inflammatory disorder; stroke; cardiovascular disease; gene therapy;
 KW vaccine; cytostatic; cerebroprotective; vulnerary; osteopathic.
 XX
 OS Homo sapiens.
 XX
 PN W0200104141-A2.
 PD 18-JAN-2001.
 XX
 PF 04-JUL-2000; 2000WO-EP06211.
 XX
 PR 12-JUL-1999; 99EP-0113428.
 XX
 PA (MERE) MERCK PATENT GMBH.
 XX
 PI Suendermann B, Hofmann U, Matzku S, Wilbert O;
 XX WPI: 2001-147177/15.
 DR N-PSDB; AAD02556.
 PT New extracellular serine protease Seripancrin, useful for treating
 PT cancer, osteoporosis, arthritis, chronic obstructive pulmonary disease,
 PT diabetes, inflammatory disorders, stroke, angiogenesis and aberrant
 PT wound healing
 PT
 PS Claim 2; Page 39-40; 45pp; English.
 XX
 CC The present invention relates to seripancrin polynucleotides, and
 CC polypeptides encoded by them. Seripancrin are members of serine
 CC protease family. This protein contains a transmembrane domain,
 CC a low density lipoprotein (LDL) domain, protease domain and a
 CC scavenger receptor cysteine-rich (SRCR) domain. The LDL and SRCR

PR	14-OCT-1998;	98US-0104927.
PR	20-OCT-1998;	98US-0104957.
PR	20-OCT-1998;	98US-0105000.
PR	20-OCT-1998;	98US-0105002.
PR	21-OCT-1998;	98US-0105104.
PR	22-OCT-1998;	98US-0105169.
PR	22-OCT-1998;	98US-0105266.
PR	26-OCT-1998;	98US-0105693.
PR	26-OCT-1998;	98US-0105694.
PR	27-OCT-1998;	98US-0105807.
PR	27-OCT-1998;	98US-0105881.
PR	27-OCT-1998;	98US-0105882.
PR	27-OCT-1998;	98US-0106062.
PR	28-OCT-1998;	98US-0106023.
PR	28-OCT-1998;	98US-0106029.
PR	28-OCT-1998;	98US-0106032.
PR	28-OCT-1998;	98US-0106032.
PR	28-OCT-1998;	98US-0106178.
PR	28-OCT-1998;	98US-0106178.
PR	29-OCT-1998;	98US-0106248.
PR	29-OCT-1998;	98US-0106384.
PR	29-OCT-1998;	98US-0108500.
PR	30-OCT-1998;	98US-0106464.
PR	03-NOV-1998;	98US-0106856.
PR	03-NOV-1998;	98US-0106902.
PR	03-NOV-1998;	98US-0106902.
PR	03-NOV-1998;	98US-0106919.
PR	03-NOV-1998;	98US-0106932.
PR	03-NOV-1998;	98US-0106934.
PR	10-NOV-1998;	98US-0107783.
PR	17-NOV-1998;	98US-0108775.
PR	17-NOV-1998;	98US-0108779.
PR	17-NOV-1998;	98US-0108787.
PR	17-NOV-1998;	98US-0108788.
PR	17-NOV-1998;	98US-0108801.
PR	17-NOV-1998;	98US-0108802.
PR	17-NOV-1998;	98US-0108807.
PR	17-NOV-1998;	98US-0108807.
PR	17-NOV-1998;	98US-0108867.
PR	17-NOV-1998;	98US-0108925.
PR	18-NOV-1998;	98US-0108848.
PR	18-NOV-1998;	98US-0108849.
PR	18-NOV-1998;	98US-0108850.
PR	18-NOV-1998;	98US-0108851.
PR	18-NOV-1998;	98US-0108852.
PR	18-NOV-1998;	98US-0108858.
PR	18-NOV-1998;	98US-0108904.
PA	(GETH) GENENTECH INC.	
PI	Baker K, Goddard A, Gurney AL, Smith V, Watanabe CK, Wood WI;	
DR	WPI; 2000-237871/20.	
DR	N-PDB; AAA37099.	
XX	New mammalian DNA sequences encoding transmembrane, receptor or	
PT	secreted PRO polypeptides, useful for screening of potential peptide or	
PT	small molecule inhibitors of the relevant receptor/Ligand interactions	
XX		
PS	Claim 12; Fig 156; 773bp; English.	
XX		
XX	AAA37022 to AAA37144 encode the new isolated human transmembrane,	
CC	receptor or secreted PRO polypeptides given in AA99340 to AA99662. The	
CC	transmembrane and receptor PRO proteins can be used for screening of	
CC	potential peptide or small molecule inhibitors of the relevant	
CC	receptor/Ligand interactions. The polypeptides and nucleotide sequences	
CC	encoding them have various industrial applications, including uses as	
CC	pharmaceutical and diagnostic agents. AAA37145 to AAA37330 represent	
CC	PCR primers and hybridisation probes used in the isolation of the PRO	
CC	polypeptides from the present invention.	
XX		
XQ	Sequence 432 AA;	

	Query Match	Similarity	24.8%	Score 674	DB 21	Length 432
Best Local	Similarity	39.1%	Pred. No.1.5e-44			
Matches 148	Conservative	55	Mismatches 132	Indels 44	Gaps 12	
OY	133	CDGVSHCEGGEDEMCVRLY--GP-----	NFLQVYSQKSMHRCODDMNENYG	181		
	111	: : :	: : :			
Db	77	cdgldcldpdeehceehcvsfpegpavavrlskdgrtllqldsaigwfacdftleala	136			
OY	182	RACACDMGKGNKF-----SSGCIYVDSGGSITPMKLTNSAGNVDIYKLYHSDACSKAVY	237			
	111	: : : : : : :	: : : : :			
Db	137	etaccqmygszravelipqddldvealtensgellmtnesg-----pclsigsly	184			
OY	238	SLRCIACGVLNLSHOSRIYGGESALPGAMPQVSLHYONVHVCGGSITPEMITYAAHC	297			
	111	: : : : : : : :	: : :			
Db	185	slhclacgksl---ktrpvvggeesvdsmpqvaslgydkqlhvcgssllldpnhvltaaic	241			
OY	298	VEKPLNPNMHTAFAGILR-QSFMFYGAQYOEKVI--SHPNYDSKTKNNDIALMKLOK	353			
	111	: : : : : :	: : :			
Db	242	frkh-tldvfvkvrsgsklgsf----psalavakilliefnmpy---pkndidatlmkqf	293			
OY	354	PLTFENDIVKPVVCLPMPGMMLQLOPEQICWISGWCATEER-KGTSEVYLNAAVLLILEPORCS	412			
	111	: : : : : :	: : :			
Db	294	pltsigvtrvplicpfdeeilpractrlwliwgfctkxngqgkmsdlldqasvavctrtica	353			
OY	413	RYVYDNLITTPAMICAGFLQGNVDSOCQSGSGGLVYTSKNNIWMLIGDTWSGSCAKARPG	472			
	111	: : : : : :	: : : : : :			
Db	354	ddayagvtrlekmmcagipegvydctcgsgdgrlmyqsdl-whvvgvlsvsgycggsprtpg	412			
OY	473	YVGNMVFETDWIYRQMRAD 491				
	111	: : : : : : : :	: : : : :			
Db	413	vytkvsaylnwlynwkae 431				

RESULT	14	
AAB87581		
ID	AAB87581	standard; Protein; 432 AA.
XX		
AC	AAB87581;	
XX		
DT	15-MAY-2001	(first entry)
XX		
DE	Human PRO1570.	
XX		
KW	Human; PRO	protein; mapping.
XX		
OS	Homo sapiens.	
XX		
PN	WO200116318-A2.	
XX		
PD	08-MAR-2001.	
XX		
PF	24-AUG-2000; 2000WO-US23328.	
XX		
PR	01-SEP-1999; 99WO-US20111.	
PR	15-SEP-1999; 99WO-US21090.	
PR	07-DEC-1999; 99US-0169495.	
PR	09-DEC-1999; 99US-0170262.	
PR	11-JAN-2000; 2000US-0175481.	
PR	18-FEB-2000; 2000WO-US04341.	
PR	18-FEB-2000; 2000WO-US04342.	
PR	22-FEB-2000; 2000WO-US04414.	
PR	01-MAR-2000; 2000WO-US05601.	
PR	03-MAR-2000; 2000US-0187202.	
PR	25-APR-2000; 2000US-0199397.	
PR	22-MAY-2000; 2000WO-US14042.	
PR	05-JUN-2000; 2000US-0209832.	
XX		
PA	(GETH)	GENENTECH INC.
XX		
PI	Eaton DL, Flivetroff E, Gerritsen ME,	Goddard A, Godowski PJ
PI	Grimaldi CJ, Gurney AL, Watanabe CK,	Wood WI;
XX		

DR WPI; 2001-183260/18.
 DR N-PSDB; AAF92113.
 PT Eighty four nucleic acids encoding PRO polypeptides, useful in
 PT molecular biology, including use as hybridization probes, and in
 PT chromosome and gene mapping.
 PS
 XX Claim 12; Fig 112; 278bp; English.
 CC The present sequence is a human PRO polypeptide (secreted and
 CC transmembrane). The PRO protein, and PRO agonists, PRO antagonists or
 CC anti-PRO antibodies are useful for preparation of a medicament useful in
 CC the treatment of a condition which is responsive to the PRO protein,
 CC agonists, antagonists or anti-PRO antibodies. The PRO protein may also be
 CC employed as molecular weight markers for protein electrophoresis. The PRO
 CC coding sequence has applications in molecular biology, including use as
 CC hybridisation probes, and in chromosome and gene mapping.
 XX
 SQ Sequence 432 AA;

Query Match 24.8%; Score 674; DB 22; Length 432;
 Best Local Similarity 39.1%; Pred. No. 1.5e-44;
 Matches 148; Conservative 55; Mismatches 132; Indels 44; Gaps 12;

```

OY 133 CDGVSHCPGEGEDENRCVRLY--GP-----NFILOYSSQKRSMPVCCODMNNYNG 181
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 77 cdgdelcprlgedeehcnvkafpegpavavrlskdrclqyldsatgmwfacfdnfta1a 136
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 182 RAACDMKGYKNNFY-----SSOGIVDDSGSTSPMKLNTSAGNVDIKKLYHSDACSSKAVY 237
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 137 etaccqmgysraveigrpqddldvveltensgelmtmnsq-----pc1sgslv 184
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 238 SLRCIACGVNLNSSROSRIYVGESALPGAMPQVSLHYQNVHVCGSITTPEMIVTAAHC 297
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 165 slhclacgksl---ktrprvyggeeasvdsrwpqvslyqdkhvcgsllldrhwltaa1c 241
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 298 VEKPLNNPMHWTAFAGILR-QSFMEYAGAYQVEKYI---SHRNVDSTKKNNDIALMKLOK 353
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 242 ftkh-tdfvfwkvragsdklgsf---pslavakllilefpmty---pkndialmk1qf 293
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 354 PLTFNDLVKPVCLPMPGMMLQPEQLCMTSGMGATEEK-GKTSEVLAANKVLLIEORONS 412
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 294 pltfsgvtrp1c1p1fdeellpactrlw1lgwftkqngkmsd1llqasvqldstc1na 353
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 413 RYVYDNLITPAMICAGFLQGNVDSGCGSGPLVYSKNNIMWLLIGDTSWGSACAKAYRPG 472
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 354 ddagygvevtekmmcsag1peggyvdtcgsdsgp1lmygsdq-whvvg1vswygcggsprcp 412
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 473 YGNAVWFTDWIYROMRAD 491
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 413 ytkvsay1nwl1ynwkae 431
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

RESULT 15
 AAB6166
 ID AAB6166 standard; protein; 432 AA.
 XX
 AC AAB6166;
 XX
 DT 02-APR-2001 (first entry)
 XX
 DE Protein of the invention #78.
 XX
 KW Secreted; transmembrane; gene therapy.
 XX
 OS Unidentified.
 XX
 MO200078961-A1.
 XX
 PD 28-DEC-2000.
 XX
 PF 18-FEB-2000; 2000MO-US04342.

XX
 DR 23-JUN-1999; 99US-0141037.
 PR 20-JUL-1999; 99US-0144758.
 PR 26-JUL-1999; 99US-0145698.
 PR 01-SEP-1999; 99MO-US20111.
 PR 29-OCT-1999; 99US-0162506.
 PR 30-NOV-1999; 99MO-US28313.
 PR 02-DEC-1999; 99MO-US28551.
 PR 16-DEC-1999; 99MO-US30095.
 PR 05-JAN-2000; 2000MO-US00219.
 PR 06-JAN-2000; 2000MO-US00376.
 XX
 PA (GETH) GENENTECH INC.
 XX
 PI Baker KP, Botstein D, Desnoyers L, Eaton DL, Ferrara N, Fong S;
 PI Gao W, Goddard A, Godowski PJ, Grimaldi CJ, Gurney AL, Hillan KJ;
 PI Pan J, Peonl NF, Roy MA, Smith V, Stewart TA, Tumas D;
 PI Watanabe CK, Williams PM, Wood WI;
 XX
 DR WPI; 2001-071395/08.

Secreted and transmembrane proteins and nucleic acids designated PRO,
 useful as hybridization probes, in chromosome and gene mapping and gene
 therapy -

Claim 1; Fig 156; 787bp; English.

The present invention relates to secreted and transmembrane proteins.
 These proteins and the DNA encoding them may be used as hybridization
 probes, in chromosome and gene mapping and in the generation of
 CC anti-sense RNA and DNA. They may also be used to generate either
 CC transgenic animals or knockout animals which are in turn useful for
 CC development and screening of therapeutically useful reagents.
 The nucleic acids may also be used in gene therapy.

Sequence 432 AA;

Query Match 24.8%; Score 674; DB 22; Length 432;
 Best Local Similarity 39.1%; Pred. No. 1.5e-44;
 Matches 148; Conservative 55; Mismatches 132; Indels 44; Gaps 12;

```

OY 133 CDGVSHCPGEGEDENRCVRLY--GP-----NFILOYSSQKRSMPVCCODMNNYNG 181
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 77 cdgdelcprlgedeehcnvkafpegpavavrlskdrclqyldsatgmwfacfdnfta1a 136
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 182 RAACDMKGYKNNFY-----SSOGIVDDSGSTSPMKLNTSAGNVDIKKLYHSDACSSKAVY 237
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 137 etaccqmgysraveigrpqddldvveltensgelmtmnsq-----pc1sgslv 184
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 238 SLRCIACGVNLNSSROSRIYVGESALPGAMPQVSLHYQNVHVCGSITTPEMIVTAAHC 297
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 165 slhclacgksl---ktrprvyggeeasvdsrwpqvslyqdkhvcgsllldrhwltaa1c 241
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 298 VEKPLNNPMHWTAFAGILR-QSFMEYAGAYQVEKYI---SHRNVDSTKKNNDIALMKLOK 353
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 242 ftkh-tdfvfwkvragsdklgsf---pslavakllilefpmty---pkndialmk1qf 293
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 354 PLTFNDLVKPVCLPMPGMMLQPEQLCMTSGMGATEEK-GKTSEVLAANKVLLIEORONS 412
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 294 pltfsgvtrp1c1p1fdeellpactrlw1lgwftkqngkmsd1llqasvqldstc1na 353
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 413 RYVYDNLITPAMICAGFLQGNVDSGCGSGPLVYSKNNIMWLLIGDTSWGSACAKAYRPG 472
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 354 ddagygvevtekmmcsag1peggyvdtcgsdsgp1lmygsdq-whvvg1vswygcggsprcp 412
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 473 YGNAVWFTDWIYROMRAD 491
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 413 ytkvsay1nwl1ynwkae 431
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

Search completed: September 26, 2001, 09:22:49

Wed Sep 26 13:56:53 2001

Job time: 104 sec

us-09-615-285-2.rag

Page 13

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 26, 2001, 09:21:05 ; Search time 25.88 Seconds

(without alignments)
1448.141 Million cell updates/sec

Title: US-09-615-285-2

Perfect score: 2717

Sequence: 1 MALNSGSPRAIGPYENHGY.....VYGNVWFTDMYRQMRADG 492

Scoring table: BIOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	660	24.3	1035	1 A43090	enteropeptidase (E
2	647	23.8	1019	1 A56318	enteropeptidase (E
3	634	23.3	1034	1 A53663	enteropeptidase (E
4	577.5	21.3	1113	2 JE0315	low-density lipopr
5	569.5	21.0	638	1 KOHUP	plasma kallikrein
6	565.5	20.8	417	1 S00845	hepsin (EC 3.4.21.
7	556	20.5	638	1 KORPL	plasma kallikrein
8	538	19.8	638	1 KOMSPL	plasma kallikrein
9	533	19.6	625	1 KFHU1	coagulation factor
10	529.5	19.5	416	1 S33777	hepsin (EC 3.4.21.
11	518.5	19.1	812	1 PLMS	plasmin (EC 3.4.21
12	501	18.4	421	1 S11674	acrosin (EC 3.4.21
13	501	18.4	1524	2 T30337	polyprotein - Atri
14	490.5	18.1	415	1 A34170	acrosin (EC 3.4.21
15	489.5	18.0	436	2 JX0172	acrosin (EC 3.4.21
16	486.5	17.9	761	2 JC5759	brain-specific ser
17	483	17.8	418	2 A37344	acrosin (EC 3.4.21
18	477	17.6	790	1 PLDG	plasmin (EC 3.4.21
19	475.5	17.5	431	2 S47538	acrosin (EC 3.4.21
20	473.5	17.4	421	2 S29599	acrosin (EC 3.4.21
21	472.5	17.4	810	2 B30848	plasmin (EC 3.4.21
22	470.5	17.3	437	2 S18407	acrosin (EC 3.4.21
23	470.5	17.3	810	2 I46260	plasmin (EC 3.4.21
24	461	17.0	343	1 A57014	prostasin (EC 3.4.
25	456.5	16.8	460	2 B61545	plasmin (EC 3.4.21
26	455	16.7	812	1 PLBO	plasmin (EC 3.4.21
27	451	16.6	2616	2 A57096	nudel protein prec
28	450.5	16.6	267	2 S40006	trypsin (EC 3.4.21
29	450	16.6	270	2 S56160	mast cell tryptase

30	449	16.5	275	2 S40005	trypsin (EC 3.4.21
31	449	16.5	655	1 A46688	hepatocyte growth
32	448	16.5	277	2 S35340	trypsin (EC 3.4.21
33	447	16.5	276	2 A38654	mast cell proteina
34	446.5	16.4	420	2 A55283	acrosin (EC 3.4.21
35	445.5	16.4	810	1 PLHU	plasmin (EC 3.4.21
36	444.5	16.4	455	2 A61545	plasmin (EC 3.4.21
37	443.5	16.3	266	2 S54146	trypsin (EC 3.4.21
38	442.5	16.3	458	1 S00657	apoptofin(a) (EC
39	441.5	16.2	274	2 JC4171	trypsin (EC 3.4.2
40	441	16.2	786	1 A47547	serine proteinase
41	439	16.2	273	2 A47246	trypsin (EC 3.4.2
42	438	16.1	274	2 S35339	trypsin (EC 3.4.21
43	437	16.1	275	2 S40007	trypsin (EC 3.4.21
44	436	16.0	242	2 S49489	trypsin (EC 3.4.21
45	436	16.0	456	1 KXBO	protein C (activat

ALIGNMENTS

RESULT 1
A43090
enteropeptidase (EC 3.4.21.9) precursor - bovine
N:Alternate names: enterokinase
C:Species: Bos primigenius taurus (cattle)
C:Date: 10-Sep-1999 #sequence,revision 10-Sep-1999 #text_change 10-Sep-1999
C:Accession: A43090; A48874; A61436
R:Kitamoto, Y.; Yuan, X.; Wu, Q.; McCourt, D.W.; Sadler, J.E.
Proc. Natl. Acad. Sci. U.S.A. 91, 7588-7592, 1994
A:Title: Enterokinase, the initiator of intestinal digestion, is a mosaic protease co
A:Reference number: A43090; MUID:94329561
A:Accession: A43090
A:Status: nucleic acid sequence not shown; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-1035 <Kir>
A:Cross-references: GB:009859; NID:9746410; PIDN:AA0026.1; PID:9746411
A:Experimental source: small intestine
R:Valle, E.R.; Rehentulla, A.; Reagle, L.A.; DiBlasio, E.A.; Ferez, C.; Grant, K.L.
J. Biol. Chem. 268, 23511-23517, 1993
A:Title: Cloning and functional expression of a cDNA encoding the catalytic subunit o
A:Reference number: A48874; MUID:94043122
A:Accession: A48874
A:Molecule type: mRNA
A:Residues: 801-1035 <LAV>
A:Cross-references: GB:L19663; NID:9416131; PIDN:AA16035.1; PID:9416132
A:Note: Parts of this sequence, including the amino end of the mature protein, were c
J. Right, A.; Janska, H.
J. Protein Chem. 10, 475-480, 1991
A:Title: The amino-terminal sequence of the catalytic subunit of bovine enterokinase.
A:Reference number: A61436; MUID:92189715
A:Accession: A61436
A:Molecule type: protein
A:Residues: 801-807,'Y',809-827 <LIG>
C:Comment: The mechanism of association with the membrane of the intestinal brush bor
embrae attachment using a signal-anchor sequence.
C:Comment: Conversion from membrane-bound to soluble forms may involve further proces
C:Complex: mature enteropeptidase is variously reported to contain two (heavy and lig
1 ifde linked
C:Function:
A:Description: cleaves propeptide from trypsinogen to produce active trypsin
A:Pathway: intestinal digestive hydrolase cascade
C:Superfamily: enteropeptidase: C1r/C1s repeat homology; LDL receptor ligand-binding
C:Keywords: glycoprotein; hydrolase; intestine; serine proteinase; transmembrane prot
F:22-38/Domain: transmembrane #status predicted <TM>
F:52-117/Product: enteropeptidase mini chain #status predicted <MC>
F:118-800/Product: enteropeptidase heavy chain #status predicted <HC>
F:199-236/Domain: LDL receptor ligand-binding repeat homology <LDL1>
F:358-520/Domain: MAM homology <MAM>
F:542-641/Domain: C1r/C1s repeat homology <C1R>
F:659-693/Domain: LDL receptor ligand-binding repeat homology <LDL2>
F:694-799/Domain: scavenger receptor cysteine-rich domain homology #status atypical <
F:801-1035/Product: enteropeptidase light chain #status predicted <LC>

F:801-1030/Domain: trypsin homology <TRY>
 F:116,147,170,194,233,263,264,404,456,486,519,550,646,698,722,741,762,864,903,965/Bindin
 F:788-912,926-942,957-972,983-1011/Disulfide bonds: #status predicted
 F:841,892,987/Active site: His, Asp, Ser #status predicted

Query Match 24.3%, Score 660; DB 1; Length 1035;
 Best Local Similarity 33.9%, Pred. No. 5.7e-40;
 Matches 150; Conservative 67; Mismatches 180; Indels 46; Gaps 12;

65 VCTQKSPSGVCTSKTKALCITLT-----CTPLVGAALAGLMLKFMKSCNSNG 117
 Db VYTGCGPNDVFSTINRMTVLFITDNLAKGCFKAFNTTGGVGLG-----IPECKEDN 663

QY 118 IECDSGTCINPMSMCDVSHCPGEGEDENRCVRLYG-----PNFIQVYSQKSKSWH 169
 Db 664 FQC-KDGEICPLVNLCDGFPHCKDGDSEAHCVRLFNGTTDSSGLVQFRIO-----SIWH 716

QY 170 PYCQDDNMENYGRACRDMGKKNNTYSSQGIYDSSGTSFPMKLTNSAGNVDIYKKLHSD 229
 Db 717 VACAMNMTQISDVCQLGLGDTG--NSSVPTFTGCGPYYVNLNTPAPNGSLI--LTPSQ 771

QY 220 ACSSKAVVSLRC--IACGVNLNSSRQ--RIYGGESALPGAMPQVSLHVNHYVCGGSII 286
 Db 772 QCLEDSLILQCNKSCGKAVLTQVSPKIVGSDSRGACMPVVALYTFDDQVCGASLY 831

QY 287 TPFWITVAHCVKPLNPNMHTAFAGILROSFMEYGAQGYE-----KVISHPNYDSKT 341
 Db 832 SRDMLVSAHCHYGNRMPSKMKAVILGLHMASNL--TSQETELRLDQIVINPNYRR 888

QY 342 KNNIDALMLKRLPTFNDLVKPYCLPFGMMIQLPDLCHISGKTEKGTSEVLNAK 401
 Db 889 KNNIDAMMLLEKKNVYTYIQLPCLPENQVFPFRGICSIAGMALIYQGSYADVLQED 948

QY 402 VLLIETQCNRRYVNDLITPRMICAGFLQGVNDSCGDSGGLPYSKNNIMWLIGDTS 461
 Db 949 VPLSNKCCQOQMPEYN-ITEMNVCAGYEGAGVDSGCGSLMCGENNRMLLAGVTSF 1007

QY 462 GSGCAKAYRPGVYGNVWFTDWI 484
 Db 1008 GYOCALPNRPGYARVPRTTEWI 1030

RESULT 2
 A:36318
 enteropeptidase (EC 3.4.21.9) precursor - human
 N:Alternate names: enterokinase
 C:Species: Homo sapiens (man)
 C:Date: 19-May-1995 #sequence_revision 09-Aug-1996 #text_change 18-Jun-1999
 C:Accession: A56318; B43090
 R:Kilameto, Y.; Veille, R.A.; Donis-Keller, H.; Sadler, J.E.
 Biochemistry 34, 4562-4568, 1995
 A:Title: cDNA sequence and chromosomal localization of human enterokinase, the proteolyt
 A:Reference number: A56318; MUID:95234679
 A:Accession: A56318
 A:Molecule type: mRNA
 A:Residues: 1-1019 <KIT>
 A:Cross-references: GB:009860; NID:9746412; PIDN:AAC50138.1; PID:9746413
 R:Kilameto, Y.; Yuan, X.; Wu, Q.; McCourt, D.W.; Sadler, J.E.
 Proc. Natl. Acad. Sci. U.S.A. 91, 7588-7592, 1994
 A:Title: Enterokinase, the initiator of intestinal digestion, is a mosaic protease compo
 A:Reference number: A43090; MUID:94325561
 A:Accession: B43090
 A:Status: nucleic acid sequence not shown
 A:Molecule type: mRNA
 A:Residues: 749-1019 <KIT>
 A:Cross-references: GB:009860
 C:Comment: The mechanism of association with the membrane of the intestinal brush border
 (noted below) or with amino-terminal myristoylation of the heavy chain.
 A:Gene: GDB:PRSS7
 A:Cross-references: GDB:384083; OMIM:226200
 A:Map position: 21q21-21q21

C:Complex: Mature enteropeptidase is variously reported to contain two (heavy and lig
 ed by a disulfide bond. Possibly, conversion from membrane-bound to soluble forms inv
 ducts.

A:Function:
 A:Description: cleaves activation peptide from trypsinogen to produce active trypsin
 A:Pathway: intestinal digestive hydrolase cascade
 C:Superfamily: enteropeptidase; C1r/C1s repeat homology; LDL receptor ligand-binding
 C:Keywords: glycoprotein; hydrolase; serine proteinase; transmembrane protein; zymoge
 F:1-784/Product: enteropeptidase heavy chain #status predicted <HCH>
 F:2-38/Domain: transmembrane #status predicted <TRM>
 F:184-221/Domain: LDL receptor ligand-binding repeat homology <LDL1>
 F:342-504/Domain: MAM homology <MAM>
 F:526-631/Domain: C1r/C1s repeat homology <C1R>
 F:643-677/Domain: LDL receptor ligand-binding repeat homology <LDL2>
 F:678-783/Domain: scavenger receptor cysteine-rich domain homology #status atypical <
 F:785-1019/Product: enteropeptidase light chain #status predicted <LCH>
 F:785-1014/Domain: trypsin homology <TRY>
 F:116,147,179,328,335,388,440,470,503,534,630,682,706,725,848,887,909,949/Binding sit
 F:772-886,810-826,910-977,941-956,967-995/Disulfide bonds: #status predicted
 F:825,876,971/Active site: His, Asp, Ser #status predicted

Query Match 23.8%, Score 647; DB 1; Length 1019;
 Best Local Similarity 35.8%, Pred. No. 4.9e-39;
 Matches 138; Conservative 67; Mismatches 134; Indels 26; Gaps 10;

QY 113 CSNSGIECDSSGTCINPMSMCDVSHCPGEGEDENRCVRLYG-----PNFIQVYSQ 164
 Db 643 CKADHFQC-KNCECVPLVNLCDGHLHCEGSDPADCVAFNFGTTNNGLVFRIO----- 696

QY 165 RKSHPVCODDNMENYGRACRDMGKKNNTYSSQGIYDSSGTSFPMKLTNSAGNVDIYK 224
 Db 697 -SIWTFACAEEMTTQISDVCQLGLGSG-NSSKPIFTSDG-PRVKLNTAP--DGLHI 750

QY 225 LYHSDACSKKAVSLRC--IACGVNLNSSR--QSRVGGESALPGAMPQVSLHVNHYV 281
 Db 751 LTPSQCCILQDSILRLQCHHKSGKRLAODITPKIVGSSNAKEGMPVVALYIYGRLLC 810

QY 282 GGSITTPFWITVAHCVKPLNPNMHTAFAGILROSFME--YGAGYQEVYSHPNYDS 339
 Db 811 GASLVSSDMLVSAHCHYGNRLPSPKMTAIIGLHMKSNLTSQETVPRILDELIVPNHYR 870

QY 340 KTKNNDIALMLKRLPTFNDLVKPYCLPFGMMIQLPDLCHISGKTEKGTSEVLNA 399
 Db 871 RKNDNDIAMMLLEKKNVYTYIQLPCLPENQVFPFRGNCISAGGVYVGGTANILQ 930

QY 400 AKVLIETQCNRRYVNDLITPRMICAGFLQGVNDSCGDSGGLPYSKNNIMWLIGDT 459
 Db 931 ADVPLSNKCCQOQMPEYN-ITEMNVCAGYEGAGVDSGCGSLMCGENNRMLLAGVT 989

QY 460 SWGSCAKAYRPGVYGNVWFTDWI 484
 Db 990 SFGYKCALPNRPGYARVARSFTTEWI 1014

RESULT 3
 A:53663
 enteropeptidase (EC 3.4.21.9) precursor - pig
 N:Alternate names: enterokinase
 C:Species: Sus scrofa domestica (domestic pig)
 C:Date: 07-Oct-1994 #sequence_revision 09-Aug-1996 #text_change 18-Jun-1999
 C:Accession: A53663
 R:Matsushima, M.; Ichinose, M.; Yahagi, N.; Kakei, N.; Tsukada, S.; Miki, K.; Kurokaw
 J. Biol. Chem. 269, 19976-19982, 1994
 A:Title: Structural characterization of porcine enteropeptidase.
 A:Reference number: A53663; MUID:94327548
 A:Accession: A53663
 A:Molecule type: mRNA
 A:Residues: 1-1034 <MAT>
 A:Cross-references: GB:D30799; NID:9505122; PIDN:BA06459.1; PID:9505123
 A:Note: parts of this sequence, including the amino ends of three chains isolated fro
 C:Comment: The mechanism of association with the membrane of the intestinal brush bor
 (noted below) or with amino-terminal myristoylation of the heavy chain.

C:Complex: Mature enteropeptidase is variously reported to contain two (heavy and light) ed by a disulfide bond. Possibly, conversion from membrane-bound to soluble forms involv
C:Function:
A:Description: cleaves activation peptide from trypsinogen to produce active trypsin
A:Pathway: intestinal digestive hydrolase cascade
C:Superfamily: enteropeptidase; Clr/Cls repeat homology; LDL receptor ligand-binding rep
C:Keywords: glycoprotein; hydrolysis; serine proteinase; transmembrane protein; zymogen
F:22-38/Domain: transmembrane #status predicted <TM>
F:52-117/Product: enteropeptidase mini chain #status predicted <MC>
F:118-799/Product: enteropeptidase heavy chain #status predicted <HC>
F:199-236/Domain: LDL receptor ligand-binding repeat homology <LDL1>
F:357-519/Domain: MAM homology <MAM>
F:541-646/Domain: Clr/Cls repeat homology <CLR>
F:658-692/Domain: LDL receptor ligand-binding repeat homology <LDL2>
F:693-798/Domain: scavenger receptor cysteine-rich domain homology #status atypical <SRC
F:800-1034/Product: enteropeptidase light chain #status predicted <LC>
F:800-1039/Domain: trypsin homology <TRY>
F:116-147,170,194,283,343,350,403,455,485,518,549,645,697,701,721,740,761,804,863,902,96
F:787-911,825-841,925-992,956-971,982-1010/Disulfide bonds: #status predicted
F:840,891,986/Active site: His, Asp, Ser #status predicted

Query Match 23.3%; Score 634; DB 1; Length 1034;
Best Local Similarity 33.3%; Pred. No. 4,4e-38;
Matches 147; Conservative 72; Mismatches 179; Indels 44; Gaps 13;

```

QY 65 VCTQKSPSGVCTSKTKALCIT---LTLG---TELVGALLAGLLMKFMKSCNSG 117
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 610 VTTGEPVEDVSTTRMTVLTNDALTKGFKANFTTGYHIG-----IPEPKEDN 662

QY 118 IECDSGTCINSMNCDGVSHCPGGEENRCVRLYG--PNTLOYSSQKSMHPYCOD 175
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 663 FQCE-NGEVLVLLNLDGSHCKSDSEAHCVRLNGTANNGLVQFRQSLMHTACAN 721

QY 176 NMENTGRACRPMGKKN-----FTSSGCIYDSDSGSTSMKLTSAANDIYKYLHSA 230
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 722 WTTQTSDDVCCQLLGLGTGNSMPFES-----SGGSPVTKNTAPNGSLI---LTASQ 771

QY 231 GSKAVVSLRC--IAGCVNLNSSROS-RIVGESALPGAMPQVSHVQNVHVCSSIT 287
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 772 CEEDSLILLCNHRKSCGKQVAGVSPKIVGNDSEHGMPVVALIYNGQLCGSLVS 831

QY 288 PEMIYTAHCVKPLNPNMHTAFAGILRQSEMFYAGYQ-----VEKVISHPNDSKTK 342
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 832 RMLVSAACHVGRNLEPSKMKAILGLHMTSL---TSPQIYTRIDELVIMPHVRRRK 888

QY 343 NNDIALMKLQKPLTRNDLVKPYCLPNCMLQPEOLCMTISGKATBEKGTSEVLNAKV 402
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 889 DSDIAMHLEFVNVYTDYIOPICLPEENGVPPGRICSIAGKVKLYOGSPADILQEADY 948

QY 403 LLIETORCSRYVDNLIRPAMICAGFLGVNDSCGSGPLVYSKNIMWILIGTSMG 462
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 949 PLISNEKCGQQMPETN-TENNMCACTGEEGIDSCGDSGGLMCLNNRMLLAGVTSFG 1007

QY 463 SGCAKAYRPGVYGNVWFTDWT 484
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 1008 YQCALPNRPGVYARVPKFTETW 1029

```

RESULT 4
J030315
Low-density lipoprotein receptor-related protein - mouse
C:Species: Mus musculus (house mouse)
C:Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 18-Aug-2000
C:Accession: J030315
R:Tomita, Y.; Kim, D.; Magoori, K.; Fujino, T.; Yamamoto, T.T.
J. Biochem. 124, 784-789, 1998
A:Title: A novel low-density lipoprotein receptor-related protein with type II membrane
A:Reference number: J030315; MUID:98429596
A:Accession: J030315
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-1113 <TOM>

A:Cross-references: DBJ:AB013874; NID:93869144; PIDN:BA434371.1; PID:93869145
C:Superfamily: trypsin homology; LDL receptor ligand-binding repeat homology
F:337-372/Domain: LDL receptor ligand-binding repeat homology <LDL1>
F:374-408/Domain: LDL receptor ligand-binding repeat homology <LDL2>
F:410-443/Domain: LDL receptor ligand-binding repeat homology <LDL3>
F:447-483/Domain: LDL receptor ligand-binding repeat homology <LDL4>
F:484-682/Domain: LDL receptor ligand-binding repeat homology <LDL5>
F:684-720/Domain: LDL receptor ligand-binding repeat homology <LDL6>
F:723-757/Domain: LDL receptor ligand-binding repeat homology <LDL7>
F:869-1097/Domain: trypsin homology <TRY>

Query Match 21.3%; Score 577.5; DB 2; Length 1113;
Best Local Similarity 33.3%; Pred. No. 5.9e-34;
Matches 134; Conservative 67; Mismatches 155; Indels 47; Gaps 14;

```

QY 111 SKCSNGIECDSSGTCINSMNCDGVSHCPGGEENRCVRLY---GPNFLOYSSQKRS 167
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 721 SFCQDNELEEC-ANHECVPRDLMDGVDCSDSDSEWGCYTLISKNGNSSLLIVHKSARE- 778

QY 168 WHPVCODDNENYGRACRDMGKKNFYSSGIVDSDSGSTSMKLTSAAND---IYKK 224
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 779 -HHVCADGWRETLSQLCKOMGLGP--SVTKILIPQEQQWRLIYPMNENNGSTLOEL 835

QY 225 LVHSDACSKAVVSLRCIA--CGVNLNSSROSRIVGESALPGAMPQVSHVQ-NVHVC 281
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 836 LVYRHSCPERSSELSLCSQDQGRRAAMKRIILGKRSIRGRMPCQSLQSEPSGHIC 895

QY 282 GGSITPEMIYTAHCV-----KPLNPNMHTAFAGILRQSEMFYAGYQ 327
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 896 GCVLLAKKVVLTVAHCFEGREDADVKKVFGIINNLDHP-----SGFMQTRF----- 941

QY 328 VEKVISHPYDSTKKNDAIMKQLKPLFNDLYKPYCLPNCMLQPEOLCMTISGKAT 387
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 942 VFTILHPRIYSAVDYDISVELSDINETSIVRPELSEPELEPTDYITGWC-- 999

QY 388 EEKGTSEVLNAKAVLLIETORCSRYVDNLITPAMICAGFLGVNDSCGSGPLVY 447
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 1000 HGNKMPKFLQGEVRIITPLEQCS-YFDMKITTNNMTCAGIESGVDCMKDSSGPLYC 1058

QY 448 SK-NNIMWILIGTSMGSGC-AKAYRPGVYGNVWFTDWTIROM 488
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 1059 ERPGQMTLEFGLTSMQSVCFSKVLGPGVYSNVSYEWGIEROI 1101

```

RESULT 5
K040UP
Plasma kallikrein (EC 3.4.21.34) precursor - human
N:Alternate names: kininogenin; plasma prekallikrein
C:Species: Homo sapiens (man)
C:Date: 13-Aug-1986 #sequence_revision 13-Aug-1986 #text_change 18-Jun-1999
C:Accession: A00921; A37939
R:Chung, D.W.; Fujikawa, K.; McKullen, B.A.; Davie, E.W.
Biochemistry 25, 2410-2417, 1986
A:Title: Human plasma prekallikrein, a zymogen to a serine protease that contains fou
A:Reference number: A00921; MUID:86243359
A:Accession: A00921
A:Molecule type: mRNA
A:Residues: 1-638 <CHD>
R:Cross-references: GB:M3143; NID:9190262; PIDN:AAA60153.1; PID:9190263
R:McMullen, B.A.; Fujikawa, K.; Davie, E.W.
Biochemistry 30, 2050-2056, 1991
A:Title: Location of the disulfide bonds in human plasma prekallikrein: the presence
A:Reference number: A37939; MUID:91152016
A:Accession: A37939
A:Molecule type: Protein
A:Residues: 20-27,40-46, 'X',48,'H',50,'X',52-70,'H',75-76,'X',78-80,103-113,131-140,1
:260-283,'X',285,287-291,'X',293-295,314-317,'X',319-320,321-324,'X',329-333,334-339,
525:538-551,562,'X',564-567,573,'X',575-576,578-583,'X',585:592-604 <MCW>
C:Comment: This protein, synthesized in the liver, circulates as a noncovalent comple
C:Comment: The zymogen is activated by factor XIIIa, which cleaves the molecule into a
are linked by one or more disulfide bonds.
C:Comment: The enzyme cleaves Lys-Arg and Arg-Ser bonds. It activates, in a reciproca

Inogen and may also play a role in the renin-angiotensin system by converting prorenin A
C:Genetics:

A:Gene: GDB:KLR3

A:Cross-references: GDB:127575; OMIM:229000

A:Map position: 4q35-4q35

C:Superfamily: coagulation factor XI; trypsin homology

C:Keywords: blood coagulation; duplication; fibrinolysis; glycoprotein; hydrolase; Infla

F:1-19/Domain: signal sequence #status predicted <SIG>

F:20-63/Domain: plasma kallikrein #status predicted <MAN>

F:20-109/Domain: plasma kallikrein heavy chain #status predicted <HCH>

F:20-109/Domain: apple repeat <AP1>

F:110-189/Domain: apple repeat <AP2>

F:200-289/Domain: apple repeat <AP3>

F:391-380/Domain: apple repeat <AP4>

F:391-638/Domain: plasma kallikrein light chain #status predicted <LCH>

F:21-104,47-77,51-57,111-194,137-166,141-147,201-284,227-256,231-237,292-375,322-328,383

F:127,308,396,453,494/Binding site: carbohydrate (asn) (covalent) #status experimental

F:318-347,340-345/Disulfide bonds: #status predicted

F:390-391/Cleavage site: Arg-Ile (coagulation factor XIIa) #status predicted

F:434,483,578/Active site: His, Asp, Ser #status predicted

Query Match 21.0%; Score 569.5; DB 1: Length 638;

Best Local Similarity 32.8%; Pred. No. 1.2e-33;

Matches 155; Conservative 64; Mismatches 160; Indels 93; Gaps 23;

OY 55 RVLFGASNPVCTQPKSPSGVCTSKTKKALCITLTLGFLVGALAGLAKFMFGS--- 111

DB 218 RVLTL--PAFVCR-----TICTYHPN---CLFETFTT-----NWKISQRYV 255

OY 112 ---KCSNSGIECDSD-----SG---TCINP-----SNMGDGVSHCPGEGDENRCVRL 151

DB 256 CLKTSSEGTSESSSPQENTISGYSLTCKRTLPKPSKSIYRVD--GGBELN----- 308

OY 152 YCPNFILOYSSQKRSNHPYCODDNNENYGRACRDMKGYKNNFYSSGGLVDDSGTS--- 208

DB 309 --VTFVKV-----NVCQE-----TCYKM--IRCFEFTYSLPEDECKEKKKC 347

OY 209 FMKLTSAGNDIYKLYKLYHSDACSSKAVSLRCIACGVN--LNSRSRITVGESEALPCA 266

DB 348 FLRLSMDSPPRI---ATGTQSSG---YSLRLCNTGDNSTVTTTSRIYVGTSMSCGE 401

OY 267 WPMOVSILHVQ---NVHVGGSITTPEMIVTAHVE--KPLNPMHMTAFAGILROSEFMFY 322

DB 402 WPMOVSILHVQKILQKRLHLCGSLIGHQWVLTAAHCFDGLPLDVMR--IYSGILNLSDTIK 459

OY 323 GAGY--QVEKVISHPYVDSKTKNDIALMLKQKPLFNDLVKRVCCPRNGMQLQPEQLCM 381

DB 460 DTFPSQIKELIITHQYKVSSEGNHDLALIKLQAPLNYTEOKPCLPSKGDSTIYTCNV 519

OY 382 SGWGATEERKGTSEVLNAKAVLLIETORCNSRYVNDLITPAMICAGFLQGNVDSGCGS 441

DB 520 TGMGSKRKEGTONLQAVNIPLVINECQKRY--QDYKITQMGVACGKREGKADCKGDS 578

OY 442 GGPLVTSNNIMWLIGDTSWSSGCAKAVRPGVYGVNVWFTDVIYRQMR-ADG 492

DB 579 GGPLVCKHNGMRLVIGITSWEGCARQPGVYTVKAEYMDIILEKTSDDG 630

RESULT 6

S00845

hepsin (EC 3.4.21.-) - human

C:Species: Homo sapiens (man)

C:Date: 31-Dec-1988 #sequence_revision 31-Dec-1988 #text_change 18-Jun-1999

C:Accession: S00845

R:Legend, S.P.: Loeb, K.R.; Hagen, F.S.; Kurachi, K.; Davie, E.W.

Biochemistry 27, 1067-1074, 1988

A:Title: A novel trypsin-like serine protease (hepsin) with a putative transmembrane dom

A:Reference number: S00845; MUID:88209431

A:Accession: S00845

A:Molecule type: mRNA

A:Residues: 1-417 <LEY>

A:Cross-references: EMBL:X07732; NID:932063; PIDN:CAA30558.1; PID:932064

C:Genetics:

A:Gene: GDB:HPN; TMPRSS1; hepsin

A:Cross-references: GDB:135685; OMIM:142440

A:Map position: 19q11-19q13.2

C:Superfamily: hepsin; trypsin homology

C:Keywords: hydrolase; liver; serine proteinase; transmembrane protein

F:23-45/Domain: transmembrane #status predicted <TMN>

F:163-400/Domain: trypsin homology #status predicted <TRY>

F:188-204,291-359,322-338,349-381/Disulfide bonds: #status predicted

F:203,257,353/Active site: His, Asp, Ser #status predicted

Query Match 20.8%; Score 565.5; DB 1: Length 417;

Best Local Similarity 30.7%; Pred. No. 1.4e-33;

Matches 133; Conservative 68; Mismatches 173; Indels 59; Gaps 9;

OY 77 CTSTKTKKALCTLTTLTGTLVGALALAGLAKFMFGSKCSNSGIECDSSGTCINPSNMGDGV 136

DB 12 CCSREKVA---ALFAGTLTLTLTAIGA-----ASMAIYA 41

OY 137 SHCPGEGDENRCVRLYGFNFILOYSSQKRSNHPYCODDNNENYGRACRDMGYKNNFY 196

DB 42 VLKNSDQELIPVOYSSADARLWFDKTEGTRLLCSSRSHARVAGLSCEDMGFLRALTH 101

OY 197 SGQIVDDSGTSFKMLNTSAGNVDTYK-KLYHSDA-----CSSKAVVSLRCIACGV 246

DB 102 SELDVRTAGA-----NCTSGFPCVDEGRDLPHQTQLLEVISVCDPDRGRFLAICQDGC- 154

OY 247 NLNRSRRIYCGESALRGANFPOYSLHVQNVHVGGSITTPEMIVTAHVEKPLNPMW 306

DB 155 -RRKLPRVIRIYGRDTSIGRPMPOVSLRYDGAHLGSLSGDWVLTAAHCFPRNRYLS 213

OY 307 HMTAFAGILROSEFMFYAGYQVEKVISHPN-----DSKTKNDIALMLKQKPLTFNDL 360

DB 214 RRRVFAAGAAQASP-HGIQLGVQAVYVHGTLFPDPNPSSENSNDIALVHSSPLPLEY 272

OY 361 VKPVCPLPNEGMMLEQOLCWSIWGATEERKGTSEVLNAKAVLLIETORCNSRYVNDL 420

DB 273 IOPVCLPAAGQALVYDKICTVYGMGNTQYVQQAQVLEARVPIISNDVNCADPFGNQI 332

OY 421 TPAMICAFILQGNVDSGCGDSGGLY-----TSKNIMWLIGDTSWSSGCAKAVRPGVYN 476

DB 333 KPKMFCAGYPEGIDACGDSGSGRPVCEDSISKPRMRLCGIVSGTCLALQAKRGVYTK 392

OY 477 VWFVTDVIYRQMR 489

DB 393 VSDPRFWLFOAIK 405

RESULT 7

KORTPL

plasma kallikrein (EC 3.4.21.34) precursor - rat

N:Alternate names: Fletcher factor; kininogenin; serum kallikrein

C:Species: Rattus norvegicus (Norway rat)

C:Date: 30-Sep-1992 #sequence_revision 30-Sep-1992 #text_change 18-Jun-1999

C:Accession: A39180; A33320; S06851; I53041; S06852

R:Beaubien, G.; Rosinski-Chaplin, I.; Mattei, M.G.; Mdkay, M.; Chretien, M.; Seldah,

Biochemistry 30, 1628-1635, 1991

A:Title: Gene structure and chromosomal localization of plasma kallikrein.

A:Reference number: A39180; MUID:91129236

A:Accession: A39180

A:Molecule type: DNA

A:Residues: 1-638 <BEA>

A:Cross-references: GB:J05315

A>Note: the authors translated the codon GAG for residue 81 as Gln

R:Seldah, N.G.; Lademellem, R.; Mdkay, M.; Hamelin, J.; Lutfaia, G.; Rougeon, F.; La

DNA 8, 563-574, 1989

A:Title: The cDNA structure of rat plasma kallikrein.

A:Reference number: A33320; MUID:90091743

A:Accession: A33320

A>Status: not compared with conceptual translation

A:Molecule type: mRNA

A:Residues: 1-338 <SE>
A:Cross-references: GB:M30282; NID:g205010; PIDD:AAA41463.1; PID:g205011
A:Note: part of this sequence, including the amino ends of both the heavy and light chain
B:Paquin, J.; Benhammet, S.; Sawyer, N.; Lazure, C.; Chretien, M.; Seidah, N.G.
B:Biochim. Biophys. Acta 999, 103-110, 1989
A:Title: Rat plasma kallikrein: purification, NH(2)-terminal sequencing and development
A:Accession number: S06851; MUID:90089457
A:Accession: S06851.
A:Molecule type: protein
A:Residues: 20-45;391-413 <PAQ>
B:Seidah, N.G.; Ladehheim, R.; Mbikay, M.; Hamelin, J.; Lutfalla, G.; Rougeon, R.; Lazure
DNA Cell Biol. 8, 563-574, 1989
A:Title: The cDNA structure of rat plasma kallikrein.
A:Reference number: I53041
A:Accession: I53041
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-638 <RES>
A:Cross-references: GB:M58590; NID:g206721; PIDD:AAA42069.1; PID:g206722
C:Comment: This protein, synthesized in the liver, circulates as a noncovalent complex w
C:Comment: The zymogen is activated by factor XIIa, which cleaves the molecule into a 11kD
are linked by one or more disulfide bonds.
C:Genetics:
A:Gene: PK
C:Superfamily: coagulation factor XI; trypsin homology
C:Keywords: blood coagulation; duplication; fibrinolysis; glycoprotein; hydrolase; infla
F:1-19/Domain: signal sequence #status predicted <SIG>
F:20-390/Product: plasma kallikrein heavy chain #status experimental <MAT1>
F:20-109/Domain: apple repeat <AP1>
F:110-199/Domain: apple repeat <AP2>
F:200-289/Domain: apple repeat <AP3>
F:291-380/Domain: apple repeat <AP4>
F:391-621/Product: plasma kallikrein light chain #status experimental <MAT2>
F:391-621/Domain: trypsin homology <TRY>
F:121-104,47-77,51-57,111-194,137-166,141-147,201-284,227-256,231-237,292-375,318-347,322-
F:127,215,308,453,459,444/Binding site: carbohydrate (Asn) (covalent) #status predicted
F:396/Binding site: carbohydrate (Asn) (covalent) #status experimental
F:434,483,578/Active site: His, Asp, Ser #status predicted

C>Date: 30-Sep-1992 #sequence_revision 30-Sep-1992 #text_change 18-Jun-1999

C/Accession: A36557

R/Setdalan, N.G.; Sawyer, N.; Hamelin, J.; Mion, P.; Beaudien, G.; Brachpapa, L.; Roche

DNA Cell Biol. 9, 737-748, 1990

A/Title: Mouse plasma kallikrein: CDNA structure, enzyme characterization, and compar

A/Reference number: A36557; MUID:91090844

A/Accession: A36557

A:Molecule type: mRNA

A/Residues: 1-638 <SEI>

A/Cross-references: GB:M58568; NID:g200358; PIDN:AAA63393.1; PID:g200359

A/Note: part of this sequence, including the amino ends of both the heavy and light c

C/Comment: This protein, synthesized in the liver, circulates as a noncovalent comple

C/Comment: The zymogen is activated by factor XIa, which cleaves the molecule into a

are linked by one or more disulfide bonds.

C/Superfamily: coagulation factor XI; trypsin homology

C/Keywords: blood coagulation; duplication; fibrinolysis; glycoprotein; hydrolase; in

F:1-19/Domain: signal sequence #status predicted <SIG>

F:20-339/Product: plasma kallikrein heavy chain #status experimental <HCH>

F:20-109/Domain: apple repeat <AP1>

F:110-189/Domain: apple repeat <AP2>

F:200-289/Domain: apple repeat <AP3>

F:291-380/Domain: apple repeat <AP4>

F:391-638/Product: plasma kallikrein light chain #status experimental <LCH>

F:391-631/Domain: trypsin homology <TRY>

F:21-104,47-77,51-57,111-194,137-166,141-147,201-284,227-256,231-237,292-375,318-347,,

F:127,215,308,396,494/Binding site: carbohydrate (asn) (covalent) #status predicted

F:434,483,578/Active site: His, Asp, Ser #status predicted

Query Match 19.88; Score 538; DB 1; Length 638;

Best Local Similarity 37.18; Pred. No. 2,2e-31;

Matches 115; Conservative 52; Mismatches 117; Indels 26; Gaps 6;

QY 185 CRDMGKNNFSSOGIVDSSGTSFPMKLTNSAGNVDIKYKLYHSDPSSKAAVSLRAC 244

Db 340 CKEBCKKSLSS---TDGSPRTITYGQSSGYSRLRCKLVDPEDCTTKI----- 387

QY 245 GVNLSNFSQSHRIVGESALPGAMPQVSLAHQVNV---HYVGSITTPMIVTAHCVCKP 301

Db 388 -----NARIYGTGNSLSEMPQVSLQYKIVSQTHLGGISITGRQVLTAAHCFD-G 438

QY 302 LNNPHNMAFAGILRQSMFYGA-GYQYEKYISHNVDYSKRRKNDIAIMLKQPLTFNDL 360

Db 439 IPRPWMIYIGGILSLSEIRTEPPSSRIKELIHQYEVYSEGNDAIILKQPLNTYEF 498

QY 361 VKPCLPRLPGMWLDPEDLCWISGMCATEEKGKTSFVLAARVLIETORCNSRYVDNL 420

Db 499 QKPLCPKADPTNTITNCWYTGKGYTEKEGFTQNIILKATILPLVNECCOKY-RDYVI 557

QY 421 TPAMICAGFLQGVNDSCGSDSGSLVTSKNNIMWLIGDTSMGSGCAKAYRPGVGNVVF 480

Db 558 NKMVICAGYKRGSGTDACGSDGSLVCKHSGRMQLVGTSMGEGGRKDGQVYTVKVS 617

QY 481 TDWITYRKMR 490

Db 618 MDWILEKTOS 627

RESULT 9

KFH01

coagulation factor XIa (EC 3.4.21.27) precursor [validated] - human

N/Alternate names: antihemophilic factor C; plasma thromboplastin antecedent

C/Species: Homo sapiens (man)

C/Date: 13-Aug-1986 #sequence_revision 26-May-1994 #text_change 08-Dec-2000

C/Accession: A27431; A00920; A37940

R/Asakai, R.; Davile, E.W.; Chung, D.W.

Biochemistry 26, 7221-7228, 1987

A/Title: Organization of the gene for human factor XI.

A/Reference number: A27431; MUID:88107663

A/Accession: A27431

A:Molecule type: DNA

A/Residues: 1-625 <ASAS>

A/Cross-references: GB:M18295

A>Note: The sequence shown follows the authors' translation
R:Fujikawa, K.; Chung, D.W.; Hendrickson, L.E.; Davie, E.W.
Biochemistry 25, 2417-2424, 1986
A>Title: Amino acid sequence of human factor XI, a blood coagulation factor with four
A:Reference number: A00920; MUID:86243360
A:Accession: A00920
A:Molecule type: mRNA
A:Residues: 1-625 <FU>
A:Cross-references: GB:M13142; NID:g182832; PIDN:MAA52487.1; PID:g182833
R:McMullen, B.A.; Fujikawa, K.; Davie, E.W.
Biochemistry 30, 2056-2060, 1991
A>Title: Location of the disulfide bonds in human coagulation factor XI: the presence of
A:Reference number: A37940; MUID:91152017
A:Accession: A37940
A:Molecule type: protein
A:Residues: 28-33,35-49, 'X', 51-55, 'X', 57-63, 70-75, 'X', 77-79, 107-109, 'X', 111-112, 132-139,
'X', 140-141, 143-144, 146-147, 149-150, 152-153, 155-156, 158-159, 161-162, 164-165, 167-168, 170-171,
'X', 173-174, 176-177, 179-180, 182-183, 185-186, 188-189, 191-192, 194-195, 197-198, 200-201, 203-204, 206-207, 209-210, 212-213, 215-216, 218-219, 221-222, 224-225, 227-228, 230-231, 233-234, 236-237, 239-240, 242-243, 245-246, 248-249, 251-252, 254-255, 257-258, 260-261, 263-264, 266-267, 269-270, 272-273, 275-276, 278-279, 281-282, 284-285, 287-288, 290-291, 293-294, 296-297, 299-300, 302-303, 305-306, 308-309, 311-312, 314-315, 317-318, 320-321, 323-324, 326-327, 329-330, 332-333, 335-336, 338-339, 341-342, 344-345, 347-348, 350-351, 353-354, 356-357, 359-360, 362-363, 365-366, 368-369, 371-372, 374-375, 377-378, 380-381, 383-384, 386-387, 389-390, 392-393, 395-396, 398-399, 401-402, 404-405, 407-408, 410-411, 413-414, 416-417, 419-420, 422-423, 425-426, 428-429, 431-432, 434-435, 437-438, 440-441, 443-444, 446-447, 449-450, 452-453, 455-456, 458-459, 461-462, 464-465, 467-468, 470-471, 473-474, 476-477, 479-480, 482-483, 485-486, 488-489, 491-492, 494-495, 497-498, 500-501, 503-504, 506-507, 509-510, 512-513, 515-516, 518-519, 521-522, 524-525, 527-528, 530-531, 533-534, 536-537, 539-540, 542-543, 545-546, 548-549, 551-552, 554-555, 557-558, 560-561, 563-564, 566-567, 569-570, 572-573, 575-576, 578-579, 581-582, 584-585, 587-588, 590-591, 593-594, 596-597, 599-600, 602-603, 605-606, 608-609, 611-612, 614-615, 617-618, 620-621, 623-624, 626-627, 629-630, 632-633, 635-636, 638-639, 641-642, 644-645, 647-648, 650-651, 653-654, 656-657, 659-660, 662-663, 665-666, 668-669, 671-672, 674-675, 677-678, 680-681, 683-684, 686-687, 689-690, 692-693, 695-696, 698-699, 701-702, 704-705, 707-708, 710-711, 713-714, 716-717, 719-720, 722-723, 725-726, 728-729, 731-732, 734-735, 737-738, 740-741, 743-744, 746-747, 749-750, 752-753, 755-756, 758-759, 761-762, 764-765, 767-768, 770-771, 773-774, 776-777, 779-780, 782-783, 785-786, 788-789, 791-792, 794-795, 797-798, 800-801, 803-804, 806-807, 809-810, 812-813, 815-816, 818-819, 821-822, 824-825, 827-828, 830-831, 833-834, 836-837, 839-840, 842-843, 845-846, 848-849, 851-852, 854-855, 857-858, 860-861, 863-864, 866-867, 869-870, 872-873, 875-876, 878-879, 881-882, 884-885, 887-888, 890-891, 893-894, 896-897, 899-900, 902-903, 905-906, 908-909, 911-912, 914-915, 917-918, 920-921, 923-924, 926-927, 929-930, 932-933, 935-936, 938-939, 941-942, 944-945, 947-948, 950-951, 953-954, 956-957, 959-960, 962-963, 965-966, 968-969, 971-972, 974-975, 977-978, 980-981, 983-984, 986-987, 989-990, 992-993, 995-996, 998-999, 1001-1002, 1004-1005, 1007-1008, 1010-1011, 1013-1014, 1016-1017, 1019-1020, 1022-1023, 1025-1026, 1028-1029, 1031-1032, 1034-1035, 1037-1038, 1040-1041, 1043-1044, 1046-1047, 1049-1050, 1052-1053, 1055-1056, 1058-1059, 1061-1062, 1064-1065, 1067-1068, 1070-1071, 1073-1074, 1076-1077, 1079-1080, 1082-1083, 1085-1086, 1088-1089, 1091-1092, 1094-1095, 1097-1098, 1100-1101, 1103-1104, 1106-1107, 1109-1110, 1112-1113, 1115-1116, 1118-1119, 1121-1122, 1124-1125, 1127-1128, 1130-1131, 1133-1134, 1136-1137, 1139-1140, 1142-1143, 1145-1146, 1148-1149, 1151-1152, 1154-1155, 1157-1158, 1160-1161, 1163-1164, 1166-1167, 1169-1170, 1172-1173, 1175-1176, 1178-1179, 1181-1182, 1184-1185, 1187-1188, 1190-1191, 1193-1194, 1196-1197, 1199-1200, 1202-1203, 1205-1206, 1208-1209, 1211-1212, 1214-1215, 1217-1218, 1220-1221, 1223-1224, 1226-1227, 1229-1230, 1232-1233, 1235-1236, 1238-1239, 1241-1242, 1244-1245, 1247-1248, 1250-1251, 1253-1254, 1256-1257, 1259-1260, 1262-1263, 1265-1266, 1268-1269, 1271-1272, 1274-1275, 1277-1278, 1280-1281, 1283-1284, 1286-1287, 1289-1290, 1292-1293, 1295-1296, 1298-1299,

```

Db      526 DKLITLTQAKRILEPVTNECCOKRY - RGKHITHKMKACAGYREGGNKQACGSDGGPLSCNNH 584
OY      451 NITWMLIGDTSMWGSCAKAYRPGVYGNAVMEFTDWTIRQMRA 490
       : | | ||| |||| | | | | | | | | | | | | | | | | |
Db      585 EVMHLVGITSWGEGCAORRRPCGVYTNVVEYVDIMLEKTOA 624

RESULT 10
S33777
hepsin (EC 3.4.21.-) - rat
C:Species: Rattus norvegicus (Norway rat)
C:Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 18-Jun-1999
C:Accession: S33777; #sequence_revision 06-Jan-1995 #text_change 18-Jun-1999
B:Farley, D.; Raymond, F.; Nick, H.
Biochim. Biophys. Acta 1173, 350-352, 1993
A>Title: Cloning and sequence analysis of rat hepsin, a cell surface serine proteinase
A:Reference number: S33777; MUID:93305733
A:Accession: S33777
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-416 <FAR>
A:Cross-references: EMBL:X70900; NID:557928; PIDN:CA50256.1; PID:957929
C:Superfamily: hepsin; trypsin homology
C:Keywords: hydrolase; liver; serine proteinase; transmembrane protein
F:12-44/Domain: transmembrane #status predicted <TM>
F:162-399/Domain: trypsin homology <TRY>
F:187-203,290-358,321-337,348-380/Dissulfide bonds: #status predicted
F:202,256,352/Active site: His, Asp, Ser #status predicted

Query Match          19.5%, Score 529.5; DB 1; Length 416;
Best Local Similarity 30.6%; Pred. No. 5.5e-31;
Matches 132; Conservative 66; Mismatches 177; Indels 57; Gaps 10;

OY      77 CTSKRKKALCIITLTGFELVGAALAAGLLMKRMKSCKSNSGEEDSSGTCLNPNSWDCGY 136
       | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      11 CSKRKA---ALTGTTL-----FL-----TGIGAASIAIVT 40

OY      137 SHCPGEDENRCVRILYPNFILQVYSQQRKNHPYCDDMNENTGRAACRDNGYKNFEYS 196
       : : | | | | | | | | | | | | | | | | | | | | | |
Db      41 ILLRDDEPLVQVLSPGDSRLLYLDKTEGRFRILLCSRSNAKVAGLGCEEGLRALAH 100

OY      197 SGGIYDDSGSTFEMKLNTSA-----GNVDIYKKLIYHSDA---CSSKAVYSLKCLAGYN 247
       : : : | | | | | | | | | | | | | | | | | | | | | |
Db      101 SELDVRTVAGAN-----GSTGFCYDEGGILPAQLRLDIVISCDCEPRGRFLPATQDGC-- 153

OY      248 LNSSROSRIYGESALPCAMPQVLSLHVONVHGCGSIITEFWITTAHCHCEKPLPNFMH 307
       ||||| : | | | | | | | | | | | | | | | | | | | | |
Db      154 RKRLPYDRIVRGQDSSLRPMWPQVSLARYDGTHLCGSLDSGDVLTAAHCPEERNRVLSR 213

OY      308 WTPAFGLIRSGMFEYAGIYOEVKVISHPNY-----DSKTKNDIALMKLOKPLTFENDLV 361
       | | | | | | | | | | | | | | | | | | | | | |
Db      214 WRVFEGAVALARTSP-HAVOLDGVADVYHGCVLPFRDPTIDENSNDIAIATHLSSSLPLEEYI 272

OY      362 KVCYLPNGGMWLQRPOLDWISMGMAETEKGSEVLNAKALLJETOCNSRYYUYNLIT 421
       : |||| : : | | | | | | | | | | | | | | | | | |
Db      273 QPVCLPPAAGQALVDKGCCTVTVGMGNTQFYGOAVVLDAPRPIAISNEYCNPSPDFTGNDIK 332

OY      422 PAMICAGELGQNVDSQGDSDSGRLV----TSKNNTMWLIGDTSMWGSCAKAYRPGVYGNV 477
       | | | | | | | | | | | | | | | | | | | | | |
Db      333 PKMFACAGYREGGIDACQDSDSGHFVCECDRIISTGSWRILCGIVSMOTGALARKRGVYTKV 392

OY      478 MYFTDWITRKMR 489
       | | | | | | | | | | | | | | | | | | | | | |
Db      393 IDRFEMIFOAIK 404

RESULT 11
PLMS
plasmn (EC 3.4.21.7) precursor - mouse
N:Contains: angiotatin; plasmnogen
C:Species: Mus musculus (house mouse)
C:Date: 20-Sep-1991 #sequence_revision 01-Nov-1996 #text_change 18-Jun-1999
```


F:43-285/Domain: trypsin homology <TRY>
 F:302-379/Region: proline-rich
 F:122-210/Binding site: carbohydrate (Asn) (covalent) #status predicted
 F:25-154/Disulfide bonds: #status predicted
 F:29-162/Disulfide bonds: #status predicted
 F:73-89/Disulfide bonds: #status predicted
 F:88-142/Disulfide bonds: #status predicted
 F:177-246/Disulfide bonds: #status predicted
 F:209-225/Disulfide bonds: #status predicted
 F:236-266/Disulfide bonds: #status predicted

Query Match 18.4%; Score 501; DB 1; Length 421;
 Best Local Similarity 40.8%; Pred. No. 6.5e-29;
 Matches 107; Conservative 42; Mismatches 87; Indels 26; Gaps 10;

OY 244 CGVNLSSNQS--RIVGESALPGAMPQVSLH-----QNVHVGSGSITPEMTVTAH 296
 DB 29 CGLRRONPGGVRIVGGAAHGMWVSLQIFRYSNHRHTCGSLNSRWLTAH 88
 OY 237 C-VEKPLNPMHTAFAGILROSEMFYQ-----AGYQ---VEKYISHPNDSKTKNDI 346
 DB 89 CEVGR--NNVHMRVLFQAK-----ITYGNKKPVKAPLDERYVEKIIHKKYNSATEGNDI 143
 OY 347 ALMKLQKPLTFLDYKPVCLPN-PCGMILQPEQLCWSGKATEEGK-KTSEVLAHAKVLL 404
 DB 144 ALVETTPPSGCRFTGPGCLPHFKAGLPKGSQSWAVGWGYIEEKAPRPSSTIMEARVDL 203
 OY 405 IETGRCNSRYVDNLITPMICAGFLGQVNSCGDSGGPLV--TSKNIMWLIDTSMG 462
 DB 204 IDLDLCNSTQMTNGRVOPLNNVACGIPVKGIDTCGDSGGLMKCKDSKESAVYVIGTSMG 263
 OY 463 SGCAKAYRQVYGNVWFETDI 484
 DB 264 VGCALAKRGRITATWPLNWI 285

RESULT 13
 T30337
 polyprotein - African clawed frog
 C:Species: Xenopus laevis (African clawed frog)
 C:Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 22-Oct-1999
 C:Accession: T30337
 R:Yang, J.C.; Lindsay, L.L.; Hedrick, J.L.
 submitted to the EMBL Data Library, March 1998
 A:Description: cDNA cloning of ovochymase, a chymotrypsin-like protease released from x
 A:Reference number: Z20829
 A:Accession: T30337
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-1524 <YAN>
 A:Cross-references: EMBL:U81290; NID:g2981640; PID:g2981641; PIDN:AAC24717.1

Query Match 18.4%; Score 501; DB 2; Length 1524;
 Best Local Similarity 33.9%; Pred. No. 3e-28;
 Matches 124; Conservative 59; Mismatches 129; Indels 54; Gaps 13;

OY 134 DGVSICPGEDRCRCVLPNFILOYVSSORKSWHPYCODMWNENYGRACRDGKNN 193
 DB 486 DAVEYDDAEKQQLARLCGTLPLPT-SSPENTMLIRFKTDMENSY-----PGFKVK 537
 OY 194 F-----YSSQGITVDDSGSTSPKMLNTSAGNVDIYKKLYHSDACSSKAVVSLRCIACG 245
 DB 538 FSEVPEKQSLP--VDDPTISM--LHPRATLADV-----CG 571
 OY 246 VNLSNR--QSRIVGESALPGAMPQVSLHQNVCVGGSTITPEMTVTAHCEKPLN 303
 DB 572 MAPMTPKMWLRIVYGESAPSPMPVOVIFELRTFHEGALISPOWILTAHCRRAA-- 629
 OY 304 NPMHTTAGILROSEMFYQAGYVEKYIS--HPNYSKTKNNIDIALMKLOKPLTFDL 360
 DB 630 EPSTVTYIAG--DHNRMLNESTEOIRIKITIRIHNTNSETIINDIALLYLEPDLADF 687

OY 361 VKPVLNPNQMWLQPEQLCWSGKATEEGKKTSEVLAHAKVLLIETGRCSRYVDNLI 420
 DB 688 VAPVCLPEPEEVLTPFASVCYVATGKNTAEDDQPLAGLQQLPLIDSLICNTSY-YSEL 746
 OY 421 TPAMICAGFLQG-NVDSQCGDSGGLV-TSKNNIMWLIGTDSKSGCAKAYRQVYGNV 478
 DB 747 TDHMLCAFPSSKERDQCGDSGGLVCONKEQPSIVGLVSWEGCGRVSXPGYTKVR 806
 OY 479 VFTDWI 484
 DB 807 LEFTWI 812

RESULT 14
 A34170
 acrosin (EC 3.4.21.10) precursor - pig
 N:Alternate names: 53k fucose-binding protein
 C:Species: Sus scrofa domestica (domestic pig)
 C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
 C:Accession: A34170; S08994; S02428; S04940; S16657; S02780; S10695; S12968
 R:Baba, T.; Kaishibara, S.; Matanabe, K.; Itoh, H.; Michikawa, Y.; Kikura, K.; Taked
 J. Biol. Chem. 264, 11920-11927, 1989
 A:Title: Activation and maturation mechanisms of boar acrosin zymogen based on the de
 A:Reference number: A34170; MUID:89308595
 A:Accession: A34170
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-415 <BAH>
 A:Cross-references: GB:S04950; NID:g164702; PIDN:AAA3131.1; PID:g164703
 R:Chehova, D.; Toepfer-Petersen, E.; Zucker, A.; Jonakova, V.
 Biol. Chem. Hoppe-Seyler 371, 317-323, 1990
 A:Title: Is spermminogen a modified proacrosin? Isolation, purification, and partial c
 A:Reference number: S08994; MUID:90253655
 A:Accession: S08994
 A:Molecule type: Protein
 A:Residues: 'X', '18', 'X', '20-25', 'X', '27-32', 'X', '34-38', 'X', '40-50 <CEC>
 R:Toepfer-Petersen, E.; Henschen, A.
 FEBS Lett. 226, 38-42, 1987
 A:Title: Acrosin shows zona and fucose binding, novel properties for a serine protein
 A:Reference number: S02428; MUID:88083633
 A:Accession: S02428
 A:Molecule type: Protein
 A:Residues: 17-32;40-55 <TOB>
 R:Adham, I.M.; Klemm, U.; Maier, W.M.; Hoyer-Fender, S.; Tsaousidou, S.; Engel, W.
 Eur. J. Biochem. 182, 563-568, 1989
 A:Title: Molecular cloning of preproacrosin and analysis of its expression pattern in
 A:Reference number: S04940; MUID:89325301
 A:Accession: S04940
 A:Molecule type: mRNA
 A:Residues: 1-7,9-210,'Q',212-216,'VT',219-346,'A',348-388,390-393,'GN',396,'LVE',399
 A:Cross-references: EMBL:X18844
 A:Note: the authors translated the codon CCT for residue 240 as Ala, GCC for residue
 R:Adham, I.M.
 submitted to the EMBL Data Library, March 1989
 A:Reference number: S16657
 A:Accession: S16657
 A:Molecule type: mRNA
 A:Residues: 1-7,9-210,'Q',212-216,'VT',219-346,'A',348-388,390-393,'GN',396,'LVE',399
 A:Cross-references: EMBL:X18844; NID:g1867; PIDN:CA432948.1; PID:g1868
 A:Note: the difference at the carboxyl end is due to a frameshift error
 R:Baba, T.; Michikawa, Y.; Kawakura, K.; Aral, Y.
 FEBS Lett. 244, 132-136, 1989
 A:Title: Activation of boar proacrosin is effected by processing at both N- and C-ter
 A:Reference number: S02780; MUID:89171246
 A:Accession: S02780
 A:Molecule type: Protein
 A:Residues: 17-69 <BA2>
 R:Toepfer-Petersen, E.; Steinberger, M.; von Eschenbach, C.E.; Zucker, A.
 FEBS Lett. 265, 51-54, 1990
 A:Title: Zona pellucida-binding of boar sperm acrosin is associated with the N-termin
 A:Reference number: S10695; MUID:90306316
 A:Accession: S10695

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd

OM protein - protein search, using sw model

Run on: September 26, 2001, 09:22:10 ; Search time 16.75 Seconds

1006.191 Million cell updates/sec

Title: US-09-615-285-2

Sequence: 1 MALNSGSPRAIGPYENHG.....VYGNVMTDWTYRQMRADG 492

Scoring table:

Gapop 10.0 , Gapext 0.5

Searched: 93435 seqs, 34255486 residues

Total number of hits satisfying chosen parameters: 93435

```
Minimum DB seq length: 0
```

Maximum DB seq length: 20000000000

Post-processing:	Minimum Match	0%
Minimum Match	100%	

Listing first 45 summaries

Database : SwissProt_39:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

SUMMARIES

Result	Score	Query	ID	Description			
No.		Match	length	DB			
1	2696	99.2	492	1	TMS2_HUMAN	O15393	homo sapien
2	2117	77.9	430	1	TMS2_MOUSE	Q9J198	mus musculus
3	875.5	32.2	454	1	TMS3_HUMAN	P57727	homo sapien
4	660	24.3	1035	1	ENTK_BOVIN	P98072	bos taurus
5	647	23.8	1019	1	ENTK_HUMAN	P98073	homo sapien
6	638	23.5	1069	1	ENTK_MOUSE	P97435	mus musculus
7	634	23.3	1034	1	ENTK_PIG	P98074	sus scrofa
8	569.5	21.0	638	1	KAL_HUMAN	P03382	homo sapien
9	565.5	20.8	417	1	HEPS_HUMAN	P03581	homo sapien
10	558.5	20.6	855	1	ST1A_HUMAN	Q9Y9VE	homo sapien
11	556	20.5	638	1	KAL_RAT	P14372	rattus norv
12	548.5	20.2	902	1	ST1L_MOUSE	P56677	mus musculus
13	538	19.8	638	1	KAL_MOUSE	P26262	mus musculus
14	536.5	19.7	416	1	HEPS_MOUSE	O35453	mus musculus
15	533	19.6	635	1	FAIL_HUMAN	P03351	homo sapien
16	529.5	19.5	416	1	HEPS_RAT	O05511	rattus norv
17	518.5	19.1	812	1	PLMN_MOUSE	P20918	mus musculus
18	501	18.4	421	1	ACRO_HUMAN	P10323	homo sapien
19	490.5	18.1	415	1	ACRO_PIG	P08001	sus scrofa
20	489.5	18.0	436	1	ACRO_MOUSE	P23378	mus musculus
21	486.5	17.9	761	1	NETR_MOUSE	O08762	mus musculus
22	477	17.6	790	1	PLMN_PIG	P06667	sus scrofa
23	475.5	17.5	431	1	ACRO_RABIT	P40838	oryctolagus
24	472.5	17.4	810	1	PLMN_MAMM	P12545	marmosa m
25	470.5	17.3	437	1	ACRO_RAT	P23993	rattus norv
26	470.5	17.3	810	1	PLMN_ENTEU	Q29485	eriaceus e
27	461	17.0	343	1	PPSS_HUMAN	O16551	homo sapien
28	456.5	16.8	333	1	PLMN_SHEEP	P81286	ovis aries
29	455	16.7	812	1	PLMN_BOVIN	P06668	bos taurus
30	452	16.6	875	1	NETR_HUMAN	P56730	homo sapien
31	451	16.6	2616	1	NDL_DROME	P98159	drosophila
32	450.5	16.6	267	1	TRIV_ANGOA	P53041	anopheles g
33	450	16.6	270	1	TRIV_MERUN	P53042	meriones un

34	TRY4	ANOXA	P35078	anopheles			
35	449	16.5	275	1	HGFA_HUMAN	P04756	homo sapien
36	448	16.5	277	1	TRY2_ANOGA	P35036	anopheles g
37	447	16.5	276	1	MCT6_MOUSE	P21885	mus musculu
38	445.5	16.4	810	1	PLMN_HUMAN	P00747	homo sapien
39	444.5	16.4	378	1	PLMN_HORSE	P350010	equus cabal
40	443	16.3	274	1	TRY1_ANOGA	P350035	anopheles g
41	442.5	16.3	458	1	APOA_HUMAN	P08519	homo sapien
42	441.5	16.2	274	1	TRYM_RAT	P50343	rattus norv
43	441	16.2	653	1	HGFA_MOUSE	Q91098	mus musculu
44	441	16.2	786	1	STUB_DROME	Q05131	drosophila
45	439	16.2	273	1	MCT7_MOUSE	Q02864	mus musculu

ALIGNMENTS

RESULT	1			
TMSS2_HUMAN				
ID	TMSS2_HUMAN	STANDARD:	PRT:	492 AA.
AC	O15393			
DT	15-JUL-1998 (Rel. 36, Created)			
DT	15-JUL-1998 (Rel. 36, Last sequence update)			
DT	01-OCT-2000 (Rel. 40, Last annotation update)			
DE	TRANSMEMBRANE PROTEASE, SERINE 2 (EC 3.4.21.-).			
GN	TMPS22.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eultheria; Primates; Carnivora; Homnidae; Homo.			
OX	NCBI_Taxid:9606;			
RA	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE:97468144; PubMed:9325052;			
RA	Paoloni-Giacobino A., Chen H., Peltsch M.C., Rossier C.,			
RA	Antonarakis S.E.;			
RT	"Cloning of the TMPS22 gene, which encodes a novel serine protease			
RT	with transmembrane, LDLRA, and SRCR domains and maps to 21q22.3.";			
RL	Genomics 44:309-320(1997).			
CC	-1- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN.			
CC	-1- TISSUE SPECIFICITY: EXPRESSED STRONGLY IN SMALL INTESTINE AND			
CC	WEEKLY IN SEVERAL OTHER TISSUES.			
CC	-1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE			
CC	TRYPsin FAMILY.			
CC	-1- SIMILARITY: CONTAINS 1 SRCR DOMAIN.			
CC	-1- SIMILARITY: CONTAINS 1 LDL-RECEPTOR CLASS A DOMAIN.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration on			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation at			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (see http://www.isb-sib.ch/announcement/).			
CC	or send an email to license@isb-sib.ch).			
CC	-----			
DR	EMBL: U75329; AAC51784.1; -.			
DR	HSP: P00763; IDPO.			
DR	MIM: 602060; -.			
DR	InterPro: IPR001254; -.			
DR	InterPro: IPR001314; -.			
DR	InterPro: IPR002172; -.			
DR	Pfam: PF00057; 1dl_recept.a; 1.			
DR	Pfam: PF00089; trypsin.1.			
DR	PRINTS: PR00722; CHYMOTRYPSIN.			
DR	PROSITE: PS00134; TRYPsin_HIS; 1.			
DR	PROSITE: PS00135; TRYPsin_SER; 1.			
DR	PROSITE: PS01209; LDLRA_1; 1.			
DR	PROSITE: PS50068; LDLRA_2; 1.			
KW	Hydrolase; Serine protease; Transmembrane; Signal-anchor.			
FT	DOMAIN 1 84			
FT	TRANSMEM 85 105			
FT	EXTRACELLULAR (POTENTIAL).			
FT	DOMAIN 106 492			
FT	DOMAIN 112 150			
FT	IDL-RECEPTOR CLASS A.			


```

FT DISULFID 788 912 INTERCHAIN (BY SIMILARITY).
FT DISULFID 826 842 BY SIMILARITY.
FT DISULFID 926 993 BY SIMILARITY.
FT DISULFID 957 972 BY SIMILARITY.
FT DISULFID 983 1011 BY SIMILARITY.
FT CARBOHYD 116 116 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 147 147 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 170 170 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 194 194 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 223 233 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 253 263 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 264 264 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 404 404 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 456 456 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 486 486 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 519 519 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 550 550 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 646 646 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 698 698 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 722 722 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 741 741 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 762 762 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 864 864 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 903 903 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 965 965 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT VARSPLIC 166 192 MISSING (IN SHORT ISOFORM).
FT CONFLICT 808 R -> Y (IN REF. 3).
SQ SEQUENCE 1035 AA; 114887 MW; E207970B08296E13 CRC64;

```

```

Query Match 24.3%; Score 660; DB 1; Length 1035;
Best Local Similarity 33.9%; Pred. No. 1.2e-41;
Matches 150; Conservative 67; Mismatches 180; Indels 46; Gaps 12;

```

```

QY 65 VCTQKPSGIVCTSKTKALCITLTL-----GTFVGAIALAGLMLKFMGSKCSNG 117
DB 611 VYTGPPVNDVSTNRMVTLFTIDMLAKQGFKAFTTGIGL-----IPECKEDN 663
QY 118 IECDSGTCINSNMCDGYSHCPGCEDEKRCVRLG-----PFTILOYSSQKSKSH 169
DB 664 FQC-KDGEICPLVNLCDGPFCKDGSDEAHCVLENGTDSGLVQFRIO-----SIWH 716
QY 170 PYCDDMNENYRAACRDGYKNNFYSSOGIVDSSGTSFMKLNTSAGVNDIYKRLYHSD 229
DB 717 VACAEMWTQIDDDVQCLGLGTG--NSSVPTFTSGGGYVNLNTPNSLI---LTPSQ 771
QY 230 ACSSAAVSLRC--IACGVNLSSROS-RIVGESALPGAMPQVSLHAVQNVHVCGSII 286
DB 772 QCLEDSLILLOCNYSKCGKRLVTOEVSPTKVGSDREGAMPVVALYFDQOVCGASLY 831
QY 287 TPBMVTAHCVKEKPLNNMHTAFAGILROSFMTFGAGIYVE-----KVISHPNDSKT 341
DB 832 SMDWLVAHCVYGRNMESKMAVGLHMASVL--TSQIETRLIDIVINPHNKR 888
QY 342 KNDIALMLKQPLFENDLVKPVCLPMPMLOPEOLCISGATGEKSEVILNAK 401
DB 889 KNDIAMHLEKAVNTDITQIPCLPEENOVFPFGICISAGMALIYGSSTADVQEDAD 948
QY 402 VLLIETQRCNSRYVYDNLITPAMICAGFLQGVNDSQDGGSGPLVTSKNINMLIGDTSM 461
DB 949 VLLIETQRCNSRYVYDNLITPAMICAGFLQGVNDSQDGGSGPLVTSKNINMLIGDTSM 1007
QY 462 GSGCAKAVRPGYGVNMTFTDWI 484
DB 1008 GYOCALPNRPGYARVPTFTDWI 1030

```

```

RESULT 5
ENTR_HUMAN STANDARD; PRT: 1019 AA.
AC P98073;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)

```

```

DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE ENTEROPEPTIDASE PRECURSOR (EC 3.4.21.9) (ENTEROKINASE).
GN PRSS7 OR ENTK.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euteria; Primates; Carnivora; Hominoidea; Homo.
OX NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Duoenum;
RX MEDLINE=95234679; PubMed=7718557;
RA Kitamoto Y., Velle R.A., Donis-Keller H., Sadler J.E.;
RT "CDNA sequence and chromosomal localization of human enteropeptidase,
the proteolytic activator of trypsinogen.";
RL Biochemistry 34:4562-4568 (1995).
RN [2]
RP SEQUENCE FROM N.A.
RA Holzinger A., Buck C., Maier E.M., Meyerhofer P.U., Roscher A.A.,
RA Sadler J.E., Hadorn H.B.;
RT "Genomic organization of the human enteropeptidase.";
RN Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=20289799; PubMed=10830953;
RA Hattori M., Fujiyama A., Taylor T.D., Watanabe H., Yada T.,
RA Park H.-S., Toyoda A., Ishii K., Totoki Y., Choi D.-K., Soeda E.,
RA Ohki M., Takagi T., Sakaki Y., Taudien S., Blechschmidt K., Polley A.,
RA Menzel U., Delabar J., Kumpf K., Lehmann R., Patterson D.,
RA Reichwald K., Rump A., Schillhabel M., Schudy A., Zimmermann W.,
RA Rosenthal A., Kudoh J., Shibuya K., Kawasaki K., Asakawa S.,
RA Shintani A., Sasaki T., Nagamine K., Mitsuyama S., Antonarakis S.E.,
RA Minoshima S., Shimizu N., Nordstiek G., Hornischer K., Brandt P.,
RA Scharfe M., Schoen O., Desario A., Reichelt J., Kauer G., Blecker H.,
RA Ramser J., Beck A., Klages S., Hennig S., Rieselmann L., Dagand E.,
RA Wehmeyer S., Borzym K., Gardiner K., Nizetic D., Francis F.,
RA Lehnach H., Reinhardt R., Yaspo M.-L.;
RT "The DNA sequence of human chromosome 2L.";
RL Nature 405:311-319 (2000).
RN [4]
RP SEQUENCE OF 749-1019 FROM N.A.
RC TISSUE=Duoenum;
RX MEDLINE=94329561; PubMed=8052624;
RA Kitamoto Y., Yuan X., Wu Q., McCourt D.W., Sadler J.E.;
RT "Enteropeptidase, the initiator of intestinal digestion, is a mosaic
protease composed of a distinctive assortment of domains.";
RL Proc. Natl. Acad. Sci. U.S.A. 91:7588-7592 (1994).
RN [5]
RP FUNCTION: RESPONSIBLE FOR INITIATING ACTIVATION OF PANCREATIC
PROTEOLYTIC PROENZYMES (TRYPSIN, CHYMOTRYPSIN AND CARBOXYPEPTIDASE
A). IT CATALYZES THE CONVERSION OF TRYPSINOGEN TO TRYPSIN WHICH IN
TURN ACTIVATES OTHER PROENZYMES INCLUDING CHYMOTRYPSINOGEN,
PROCARBOXYPEPTIDASES, AND PROELASTASES.
CC -1- CATALYTIC ACTIVITY: SELECTIVE CLEAVAGE OF 6-LYS-1-ILE-7 BOND IN
TRYPSINOGEN.
CC -1- SUBUNIT: HETERODIMER OF A CATALYTIC (LIGHT) CHAIN AND A
MULTIDOMAIN (HEAVY) CHAIN LINKED BY A DISULFIDE BOND.
CC -1- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN (PROBABLE).
CC -1- TISSUE SPECIFICITY: INTESTINAL BRUSH BORDER.
CC -1- PTM: THE CHAINS ARE DERIVED FROM A SINGLE PRECURSOR THAT IS
CLEAVED BY A TRYPSIN-LIKE PROTEASE.
CC -1- DISEASE: DEFECTS IN PRSS7 CAUSE LIFE-THREATENING INTESTINAL
MALABSORPTION CHARACTERIZED BY DIARRHEA AND FAILURE TO THRIVE.
CC -1- SIMILARITY: CONTAINS 2 LDL-RECEPTOR CLASS A DOMAINS.
CC -1- SIMILARITY: CONTAINS 2 CUB DOMAINS.
CC -1- SIMILARITY: CONTAINS 1 SRCR DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 MAM DOMAIN.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
TRYPSIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial

```

CC entities requires a license agreement (See <http://www.lsb-sib.ch/announce/> or send an email to license@lsb-sib.ch).

DR EMBL: U09860: AAC50138.1; -

DR EMBL: Y19124: CAB65555.1; -

DR EMBL: Y19125: CAB65555.1; JOINED.

DR EMBL: Y19126: CAB65555.1; JOINED.

DR EMBL: Y19127: CAB65555.1; JOINED.

DR EMBL: Y19128: CAB65555.1; JOINED.

DR EMBL: Y19129: CAB65555.1; JOINED.

DR EMBL: Y19130: CAB65555.1; JOINED.

DR EMBL: Y19131: CAB65555.1; JOINED.

DR EMBL: Y19132: CAB65555.1; JOINED.

DR EMBL: Y19133: CAB65555.1; JOINED.

DR EMBL: Y19134: CAB65555.1; JOINED.

DR EMBL: Y19135: CAB65555.1; JOINED.

DR EMBL: Y19136: CAB65555.1; JOINED.

DR EMBL: Y19137: CAB65555.1; JOINED.

DR EMBL: Y19138: CAB65555.1; JOINED.

DR EMBL: Y19139: CAB65555.1; JOINED.

DR EMBL: Y19140: CAB65555.1; JOINED.

DR EMBL: Y19141: CAB65555.1; JOINED.

DR EMBL: Y19142: CAB65555.1; JOINED.

DR EMBL: Y19143: CAB65555.1; JOINED.

DR EMBL: AL163218: CAB90382.1; -

DR HSSP: P00763: IDPO.

DR MEROPS: S01.156; -

DR MIM: 226200; -

DR InterPro: IPR000082; -

DR InterPro: IPR000859; -

DR InterPro: IPR000998; -

DR InterPro: IPR001190; -

DR InterPro: IPR001254; -

DR InterPro: IPR001314; -

DR InterPro: IPR002172; -

DR Pfam: PF00431: CUB; 2.

DR Pfam: PF00629: MAM; 1.

DR Pfam: PF01390: SEA; 1.

DR Pfam: PF00530: SRCR; 1.

DR Pfam: PF00057: Idl_recept_a; 2.

DR Pfam: PF00089: trypsin; 1.

DR PRINTS: PR00722: CHYMOTRYPSIN.

DR PROSITE: PS00134: TRYPSIN_HIS; 1.

DR PROSITE: PS00135: TRYPSIN_SER; 1.

DR PROSITE: PS00180: CUB; 2.

DR PROSITE: PS00740: MAM_1; 1.

DR PROSITE: PS00740: MAM_2; 1.

DR PROSITE: PS01209: LDLR_1; 2.

DR PROSITE: PS01209: LDLR_2; 2.

DR Signal-anchor: Glycoprotein; Myristate; Hydroxylase; Serine protease; Zymogen; Transmembrane; Repeat.

KW Setline protease: Zymogen; Transmembrane; Repeat.

FT CHAIN 1 785 1019 CATALYTIC CHAIN (HEAVY CHAIN).

FT TRANSMEM 19 47 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN).

FT DOMAIN 182 223 LDL-RECEPTOR CLASS A 1.

FT DOMAIN 225 334 CUB.

FT DOMAIN 342 504 MAM.

FT DOMAIN 524 634 CUB.

FT DOMAIN 641 679 LDL-RECEPTOR CLASS A 2.

FT DOMAIN 678 771 SRCR.

FT ACT_SITE 825 825 CHARGE RELAY SYSTEM (BY SIMILARITY).

FT ACT_SITE 876 876 CHARGE RELAY SYSTEM (BY SIMILARITY).

FT ACT_SITE 971 971 MYRISTATE (POTENTIAL).

FT LIPID 2 2 BY SIMILARITY.

FT DISULFID 184 197 BY SIMILARITY.

FT DISULFID 191 210 BY SIMILARITY.

FT DISULFID 204 221 BY SIMILARITY.

FT DISULFID 643 655 BY SIMILARITY.

FT DISULFID 650 668 BY SIMILARITY.

FT DISULFID 662 677 BY SIMILARITY.

FT DISULFID 772 896 INTERCHAIN (BY SIMILARITY).

FT DISULFID 810 826 BY SIMILARITY.

FT DISULFID 910 977 BY SIMILARITY.

FT DISULFID 941 956 BY SIMILARITY.

FT DISULFID 967 995 BY SIMILARITY.

FT CARBOHYD 116 116 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 147 147 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 179 179 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 328 328 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 335 335 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 388 388 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 440 440 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 470 470 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 503 503 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 534 534 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 630 630 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 682 682 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 706 706 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 725 725 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 848 848 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 887 887 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 909 909 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 949 949 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CONFLICT 134 134 O -> E (IN REF. 3).

FT CONFLICT 732 732 S -> P (IN REF. 3).

FT CONFLICT 754 771 SOQCLDLSITLQCNHNS -> RRNAKETDALSPTILIA (IN REF. 3).

FT SEQUENCE 1019 AA; 112923 MW; B6AAA245F6D4A563 CRC64;

Query Match 23.8%; Score 647; DB 1; Length 1019;

Best Local Similarity 35.8%; Pred. No. 1,1e-40;

Matches 138; Conservative 67; Mismatches 154; Indels 26; Gaps 10;

QY 113 CSNSGIECDSSGTCINPNSMCDGVSCHCPGDENECRVLYG-----PNTIQQVSSQ 164

DB 643 CKADHFGC-KNGECVPLVNLCDGHLCEDSGDEADCVAFNFGTTNNNGLVAFRIQ----- 696

QY 165 RKSHPVQCDDMNENYGAACRDMGYNKNNFYSSQCIYDSSGTSFMKINTSAGANDIYK 224

DB 697 -SIWTTACAEWMTQOISNDVOLLGLGSG-NSSKPIFSTDCG-PFVKLNTAP---DGHIL 750

QY 225 LYHSDACSSKAVVSLRC--IACGVNLNSSR-QSRVYGESALPGAMPQVSLHYQNVHC 281

DB 751 LTPSQOCLQDLSITLQCNHNSGKGLAODITPKTVGSSNMEKGMVWVGLYGGRLIC 810

QY 282 GGSITTPEMIVTAHCVKPLNNPMTAFAGILROSEFM--YGAGYQVEYVISHPNYDS 339

DB 811 GASIVSSDMVISAACHCYGRNLEPSKWTALLGLHKKSLTSPQVPRIDIVINPHNR 870

QY 340 KTKNNDIALMKLQKPLFTNDLVKPYCLDPNPGMIQPEOLCMIWISGKATSEKGTSEVINA 399

DB 871 RRKNDIAMHLEFKVNTDYIOPICLPEENQVFPFGHNCISIAGGVVYOGTTANILQE 930

QY 400 AKVILLIETQRCNSRVYVNLITPAMICGLQGNVDSQSGSGPLVYSKNNIMLLGDT 459

DB 931 ADVPLLSNERCQOQPEYN-ITENNICAGYEGGIDSCQSGSGPLKQENNRFLAGVT 989

QY 460 SWGSGCAKARPGVYGVNMFETDWT 484

DB 990 SFGYCALPNRPQVYARVSRFTWT 1014

RESULT 6

ENTR_MOUSE STANDARD; PRT; 1069 AA.

AC P97435:

DT 01-NOV-1987 (Rel. 35, Created)

DT 01-NOV-1987 (Rel. 35, Last sequence update)

DT 01-OCT-2000 (Rel. 40, Last annotation update)

DE ENTEROPEPTIDASE (EC 3.4.21.9) (ENTEROKINASE).

GN PRSS7 OR ENTK.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

NCBI_TaxID=10090;
 [1]
 SEQUENCE FROM N.A.
 STRAIN=C57BL/6; TISSUE=Duodenum;
 MEDLINE=98147142; PubMed=9486188;
 RA Yuan X., Zheng X., Lu D., Rubin D.C., Pung C.Y.M., Sadler J.E.;
 RT "Structure of murine enterokinase (enteropeptidase) and expression in
 small intestine during development.";
 RL Am. J. Physiol. 274.G342-G349(1998).
 CC - FUNCTION: RESPONSIBLE FOR INITIATING ACTIVATION OF PANCREATIC
 PROTEOLYTIC PROENZYMES (TRYPSIN, CHYMOTRYPSIN AND CARBOXYPEPTIDASE
 A). IT CATALYZES THE CONVERSION OF TRYPSINOGEN TO TRYPSIN WHICH IN
 TURN ACTIVATES OTHER PROENZYMES INCLUDING CHYMOTRYPSINOGEN,
 PROCARBOXYPEPTIDASES, AND PROELASTASES (BY SIMILARITY).
 CC - CATALYTIC ACTIVITY: SELECTIVE CLEAVAGE OF 6-LYS-1-ILE-7 BOND IN
 TRYPSINOGEN.
 CC - SUBUNIT: HETERODIMER OF A CATALYTIC (LIGHT) CHAIN AND A
 MULTIDOMAIN (HEAVY) CHAIN LINKED BY A DISULFIDE BOND (BY
 SIMILARITY).
 CC - SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN (PROBABLE).
 CC - PTM: THE CHAINS ARE DERIVED FROM A SINGLE PRECURSOR THAT IS
 CLEAVED BY A TRYPSIN-LIKE PROTEASE (BY SIMILARITY).
 CC - SIMILARITY: CONTAINS 2 LDL-RECEPTOR CLASS A DOMAINS.
 CC - SIMILARITY: CONTAINS 2 CUB DOMAINS.
 CC - SIMILARITY: CONTAINS 1 SRCR DOMAIN.
 CC - SIMILARITY: CONTAINS 1 MAM DOMAIN.
 CC - SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
 TRYPSIN FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: U73378; AAB37317.1; -
 DR HSSP: P00763; LDPO.
 DR MEROPS: S01.156; -
 DR MGD: MGI:1197523; Prss7.
 DR InterPro: IPR000082; -
 DR InterPro: IPR000859; -
 DR InterPro: IPR000998; -
 DR InterPro: IPR001190; -
 DR InterPro: IPR001254; -
 DR InterPro: IPR001314; -
 DR InterPro: IPR002172; -
 DR Pfam: PF00431; CUB; 2.
 DR Pfam: PF00629; MAM; 1.
 DR Pfam: PF01390; SEA; 1.
 DR Pfam: PF00530; SRCR; 1.
 DR Pfam: PF00057; 1dl_recept_a; 2.
 DR Pfam: PF00089; trypsin; 1.
 DR PRINTS: PR00722; CHYMOTRYPSIN.
 DR PROSITE: PS00134; TRYPSIN_HIS; 1.
 DR PROSITE: PS00135; TRYPSIN_SER; 1.
 DR PROSITE: PS01180; CUB; 2.
 DR PROSITE: PS00740; MAM_1; 1.
 DR PROSITE: PS50060; MAM_2; 1.
 DR PROSITE: PS01209; LDLRA_1; 2.
 DR PROSITE: PS50068; LDLRA_2; 2.
 KM Signal-anchor: Glycoprotein; Myristate; Hydrolase;
 KW Serine protease; Zymogen; Transmembrane; Repeat.
 FT CHAIN 1 829
 FT CHAIN 1 829
 FT TRANSMEM 19 47
 FT DOMAIN 227 268
 FT DOMAIN 270 379
 FT DOMAIN 387 549
 FT DOMAIN 569 679
 FT DOMAIN 686 724
 FT DOMAIN 723 816
 FT SRCR.

FT	ACT_SITE	874	874	CHARGE RELAY SYSTEM (BY SIMILARITY).
FT	ACT_SITE	925	925	CHARGE RELAY SYSTEM (BY SIMILARITY).
FT	ACT_SITE	1021	1021	CHARGE RELAY SYSTEM (BY SIMILARITY).
FT	LIPID	2	2	MYRISTATE (POTENTIAL).
FT	DISULFID	229	242	BY SIMILARITY.
FT	DISULFID	236	255	BY SIMILARITY.
FT	DISULFID	249	266	BY SIMILARITY.
FT	DISULFID	688	700	BY SIMILARITY.
FT	DISULFID	695	713	BY SIMILARITY.
FT	DISULFID	707	722	BY SIMILARITY.
FT	DISULFID	817	945	INTERCHAIN (BY SIMILARITY).
FT	DISULFID	859	875	BY SIMILARITY.
FT	DISULFID	959	1027	BY SIMILARITY.
FT	DISULFID	991	1006	BY SIMILARITY.
FT	DISULFID	1017	1045	BY SIMILARITY.
FT	CARBOHYD	147	147	N-LINKED (GLCNAc. . .) (POTENTIAL).
FT	CARBOHYD	197	197	N-LINKED (GLCNAc. . .) (POTENTIAL).
FT	CARBOHYD	212	212	N-LINKED (GLCNAc. . .) (POTENTIAL).
FT	CARBOHYD	373	373	N-LINKED (GLCNAc. . .) (POTENTIAL).
FT	CARBOHYD	380	380	N-LINKED (GLCNAc. . .) (POTENTIAL).
FT	CARBOHYD	433	433	N-LINKED (GLCNAc. . .) (POTENTIAL).
FT	CARBOHYD	515	515	N-LINKED (GLCNAc. . .) (POTENTIAL).
FT	CARBOHYD	579	579	N-LINKED (GLCNAc. . .) (POTENTIAL).
FT	CARBOHYD	675	675	N-LINKED (GLCNAc. . .) (POTENTIAL).
FT	CARBOHYD	727	727	N-LINKED (GLCNAc. . .) (POTENTIAL).
FT	CARBOHYD	751	751	N-LINKED (GLCNAc. . .) (POTENTIAL).
FT	CARBOHYD	770	770	N-LINKED (GLCNAc. . .) (POTENTIAL).
FT	CARBOHYD	791	791	N-LINKED (GLCNAc. . .) (POTENTIAL).
FT	CARBOHYD	897	897	N-LINKED (GLCNAc. . .) (POTENTIAL).
FT	CARBOHYD	936	936	N-LINKED (GLCNAc. . .) (POTENTIAL).
FT	CARBOHYD	999	999	N-LINKED (GLCNAc. . .) (POTENTIAL).
SO	SEQUENCE	1069 AA;	118735 MM;	E62549EA63743C3D CRC64;

Query Match 23.5%; Score 638; DB 1; Length 1069;
 Best Local Similarity 30.5%; Pred. No. 5.3e-40;
 Matches 164; Conservative 92; Mismatches 196; Indels 86; Gaps 21;

QY	18	HGYQPNPYPAPQPTVPT-----YEV-HQAQYPPSPVQYARVLTQAS----	61
DB	544	NGICSGPPY-EPTLVPTPELPDGGPFELMEBNSTFS--PMPKPYQNASCIWN	600
QY	62	-----NPV-----VCIQPKSPSGTVCCKTKAL	85
DB	601	LANQRCKNTQIHRQEDDENINDVYEVROGGERDLSLLAVYIGP-GPVADISTTRKMTV	659
QY	86	CITLLGLTFLVG--AALAAGLIMKFMG--SKCSNSGIECDSSGTCINPNSMCDGVSHPG	141
DB	660	IFTTNNETRRKGRKANFTSGY--YLGIPEPCDDDFQC-KGNCICPLNLDSDYHCHD	715
QY	142	GEDENRCVRLY----GPNFLQYYSQKRSWHPVCODDMNENTGRACAKDMGKKNFTSS	197
DB	716	GSDASCVRFELNCTRSNNGLVDF--NIHSIWHIACENMTTQISNEVCHLIGLS--ANS	771
QY	198	QGVLDVDSGSTSPFKLNTSAGNVDIYKKLYHSDACSKAVVSLRC--IACGVNLNSRSOS-	254
DB	772	SMFISSTGGGPFYRVQAPRGLI---LTPSLQCSODSLILQCHNKSCGEEKVYTKVSP	828
QY	255	RIVGSEALPGAMPQVSLHYOVNH---VCGGSITTPEMITTAACHVEKPLNPNWHTA	310
DB	829	KIYVGSADAQAGAMPVVALYHNRDSTRLLCGASVLSMVLVAACHVYRRLNDPTRMFA	888
QY	311	FAGILQSFMEFYAGYQ--VEKYISHPNVDSKTKNNDIALMKLQKPLTNDLYKPCLEPN	368
DB	889	VGLHMQSNLTPDQVVRVAVDQIVINPHYDRRKRVNDIAMHLEFVYVNTDVIQPCLE	948
QY	369	PGMILQPEOLCMTISWGATE-EKGTSEVLNNAKVLLIFETORNSRYVDNITPMMA	427
DB	949	ENQIFLIPGRKCSLAGNGYDKINAGSTVDVLEKADVPLINSEKQOQDLPEYN-ITTESMICA	1007
QY	428	GLQGNVDSQCGDSGGLVTSKNNIMWLIGTWSGSCAKAYRPGVYGVNVMFTDMY	485
DB	1008	GYEGGIDSCQGSQGLPMCKQENNRWFLVGVTSFVQCALPNHPRGVYVAVSQEIFEWH	1065

```

RESULT 7
ENTK_PIG ID ENTK_PIG STANDARD: PRT: 1034 AA.
AC P98074;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE ENTEROPEPTIDASE PRECURSOR (EC 3.4.21.9) (ENTEROKINASE).
GN PRS87 OR ENTK.
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX NCBI_TaxID=9823;
RN [1]
RP TISSUE=FROM N.A., AND PARTIAL SEQUENCE.
RC TISSUE=Duodenal mucosa;
RX MEDLINE=94327548; PubMed=8051081;
RA Matsushima M., Ichinose M., Yahagi N., Kakei N., Tsukada S.,
RA Miki K., Kurokawa K., Tashiro K., Shiohara K., Shinomiya K.,
RA Umemura H., Inoue H., Takahashi T., Takahashi K.;
RT "Structural characterization of porcine enteropeptidase.";
RL J. Biol. Chem. 269:19976-19982(1994)
CC -1- FUNCTION: RESPONSIBLE FOR INITIATING ACTIVITY OF PANCREATIC
CC PROTEOLYTIC PROENZYMES (TRYPsin, CHYMOTRYPSIN AND CARBOXYPEPTIDASE
CC A). IT CATALYZES THE CONVERSION OF TRYPSINOGEN TO TRYPSIN WHICH IN
CC TURN ACTIVATES OTHER PROENZYMES INCLUDING CHYMOTRYPSINOGEN,
CC PROCARBOXYPEPTIDASES, AND PROELASTASES.
CC -1- CATALYTIC ACTIVITY: SELECTIVE CLEAVAGE OF 6-LYS-|-ILE-7 BOND IN
CC TRYPsinOGEN.
CC -1- SUBUNIT: HETERODIMER OF A CATALYTIC (LIGHT) CHAIN, A MULTIDOMAIN
CC (HEAVY) CHAIN, AND A MINI CHAIN.
CC -1- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN (PROBABLE).
CC -1- PTM: THE CHAINS ARE DERIVED FROM A SINGLE PRECURSOR THAT IS
CC CLEAVED BY A TRYPSIN-LIKE PROTEASE.
CC -1- PTM: THE MINI CHAIN MAY BE CLEAVED BY ELASTASE.
CC -1- SIMILARITY: CONTAINS 2 LDI-RECEPTOR CLASS A DOMAINS.
CC -1- SIMILARITY: CONTAINS 2 CUB DOMAINS.
CC -1- SIMILARITY: CONTAINS 1 SCR DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 MAM DOMAIN.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1, ALSO KNOWN AS THE
CC TRYPsin FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL, D30799; BA06459.1; -.
DR HSSP, P00763; IDPO.
DR MEROPS, S01.156; -.
DR InterPro: IPR000082; -.
DR InterPro: IPR000089; -.
DR InterPro: IPR000098; -.
DR InterPro: IPR001190; -.
DR InterPro: IPR001254; -.
DR InterPro: IPR001314; -.
DR InterPro: IPR002172; -.
DR Pfam: PF00431; CUB; 2.
DR Pfam: PF00629; MAM; 1.
DR Pfam: PF01390; SEA; 1.
DR Pfam: PF00530; SKCR; 1.
DR Pfam: PF00057; Id1_recept_a; 2.
DR Pfam: PF00089; trypsin; 1.
DR PRINTS: PR00722; CHYMOTRYPsin.
DR PROSITE: PS00134; TRYPSIN_HIS; 1.
DR PROSITE: PS00135; TRYPSIN_SER; 1.
DR PROSITE: PS0180; CUB; 2.
DR PROSITE: PS00740; MAM_1; 1.

```

DR	PROSITE: P550060; MAM_2; 1.
DR	PROSITE: PS01209; LIDRA_1; 2.
DR	PROSITE: PS50068; LIDRA_2; 2.
KW	Signal-anchor; Glycoprotein; Myristate; Hydrolase;
KM	Serine protease; Zymogen; Transmembrane; Repeat.
FT	CHAIN 52 117
FT	CHAIN 118 799
FT	CHAIN 800 1034
FT	CHAIN 19 47
FT	TRANSMEM 197 228
FT	DOMAIN 240 349
FT	DOMAIN 357 519
FT	DOMAIN 539 649
FT	DOMAIN 656 694
FT	DOMAIN 693 786
FT	ACT_SITE 840 840
FT	ACT_SITE 891 891
FT	ACT_SITE 966 986
FT	ACT_SITE 2 2
FT	LIPID 2 2
FT	DISULFID 199 212
FT	DISULFID 206 225
FT	DISULFID 219 236
FT	DISULFID 658 670
FT	DISULFID 665 683
FT	DISULFID 677 692
FT	DISULFID 787 911
FT	DISULFID 825 841
FT	DISULFID 925 992
FT	DISULFID 966 971
FT	DISULFID 982 1010
FT	CARBOHYD 116 116
FT	CARBOHYD 147 147
FT	CARBOHYD 170 170
FT	CARBOHYD 194 194
FT	CARBOHYD 283 283
FT	CARBOHYD 343 343
FT	CARBOHYD 350 350
FT	CARBOHYD 403 403
FT	CARBOHYD 455 455
FT	CARBOHYD 485 485
FT	CARBOHYD 518 518
FT	CARBOHYD 549 549
FT	CARBOHYD 645 645
FT	CARBOHYD 697 697
FT	CARBOHYD 701 701
FT	CARBOHYD 721 721
FT	CARBOHYD 740 740
FT	CARBOHYD 761 761
FT	CARBOHYD 804 804
FT	CARBOHYD 863 863
FT	CARBOHYD 902 902
FT	CARBOHYD 964 964
FT	CARBOHYD 1034 AA; 114776 MW; 038BC64CF64C368 CRC64;
SEQ	SEQUENCE

```

Db 772 CREDLSILLQCNHKSCKGKROVAOEVSPKIVGNDSEHGAMPVVALYNGOLICGASLVS 831
OY 288 PEMIYTAACHVEKPLNPNHMTAFAGILRQSEMFYAGYO-----VEKISHPNVDSKTR 342
Db 832 RQMLVSAACHVGRNLEPKRKAILGLHMTSNL---TSPQIYTRLDELIVINHTRRRK 888
OY 343 NNDIALMLKQKPLTFENDLVKPYCLPNPMMLOPEOLCMTISGNGATEKCTSEVLNAAY 402
Db 889 DSDIAMMHLEFVNTYDIQPICLPEENVFPFGRICSLAGNGKVIYQSPADILQEAAY 948
OY 403 LLIETORCSRYVNDLITPAMICAGFLOGVNDSCOGSGGLVYSKNNIMWILIGDTSWG 462
Db 949 PLTNSKCCQOMPEYV-ITENNMACAGEEGGIDSCGSGGLMCLNNRMLAGVTSFG 1007
OY 463 SGCAKAYRPGVYGNVVFETDWI 484
Db 1008 YOCALPNRPGVYARVAFKTEWI 1029

RESULT 8
KAL_HUMAN STANDARD: PRT: 638 AA.
AC P03952.
DT 23-OCT-1986 (Rel. 02, Created)
DT 23-OCT-1986 (Rel. 02, Last sequence update).
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE PLASMA KALLIKREIN PRECURSOR (EC 3.4.21.34) (PLASMA PREKALLIKREIN)
DE (KININOGENIN) (FLETCHER FACTOR).
GN KLK3.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=86243359; PubMed=3521732;
RX Chung D.W., Fujikawa K., McMullen B.A., Davie E.W.;
RA "Human plasma prekallikrein, a zymogen to a serine protease that
RT contains four tandem repeats."
RL Biochemistry 25:2410-2417(1986).
RP PARTIAL SEQUENCE, AND DISULFIDE BONDS.
RX MEDLINE=91152016; PubMed=1998666;
RA McMullen B.A., Fujikawa K., Davie E.W.;
RT "Location of the disulfide bonds in human plasma prekallikrein: the
RT presence of four novel apple domains in the amino-terminal portion of
RT the molecule."
RL Biochemistry 30:2050-2056(1991).
CC -1- FUNCTION: THE ENZYME CLEAVES LYS-ARG AND ARG-SER BONDS. IT
CC ACTIVATES, IN A RECIPROCAL REACTION, FACTOR XII AFTER ITS BINDING
CC TO A NEGATIVELY CHARGED SURFACE. IT ALSO RELEASES BRADYKININ FROM
CC HMW KININOGEN AND MAY ALSO PLAY A ROLE IN THE RENIN-ANGIOTENSIN
CC SYSTEM BY CONVERTING PRORENIN INTO RENIN.
CC -1- SUBUNIT: THE ZYMOGEN IS ACTIVATED BY FACTOR XIIa, WHICH CLEAVES
CC THE MOLECULE INTO A LIGHT CHAIN, WHICH CONTAINS THE ACTIVE SITE,
CC AND A HEAVY CHAIN, WHICH ASSOCIATES WITH HMW KININOGEN. THESE
CC CHAINS ARE LINKED BY ONE OR MORE DISULFIDE BONDS.
CC -1- DISEASE: DEFECTS IN KLK3 ARE THE CAUSE OF FLETCHER FACTOR
CC DEFICIENCY, A BLOOD COAGULATION DEFECT.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC TRYPSIN FAMILY. PLASMA KALLIKREIN SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-stb.ch/announce/
CC or send an email to license@isb-stb.ch).
CC -----
CC EMBL, M13143; AAA60153.1; -
CC DR PIR: A00921; KOHUP.

```

```

DR PIR: A37939; A37939.
DR HSSP: P00763; IDPO.
DR MEROPS: S01.212; -.
DR MIM: 229000; -.
DR InterPro: IPR00177; -.
DR InterPro: IPR001254; -.
DR InterPro: IPR001314; -.
DR InterPro: IPR003014; -.
DR Pfam: PR00024; PAN. 4.
DR Pfam: PR00089; trypsin. 1.
DR PRINTS: PR00005; APPLEDOMAIN.
DR PRINTS: PR00722; CHYMOTRYPSIN.
DR PROSITE: PS00134; TRYPSIN_SER. 1.
DR PROSITE: PS00135; TRYPSIN_SER. 1.
DR PROSITE: PS00495; APPLE. 4.
KW Hydrolyase; Serine protease; Glycoprotein; Plasma; Zymogen; Signal;
KW Fibrinolysis; Blood coagulation; Inflammatory response; Liver;
KW Repeat.
FT SIGNAL 1 19
FT CHAIN 20 390 PLASMA KALLIKREIN HEAVY CHAIN.
FT CHAIN 391 638 PLASMA KALLIKREIN LIGHT CHAIN.
FT DOMAIN 20 105 APPLE 1.
FT DOMAIN 110 195 APPLE 2.
FT DOMAIN 200 285 APPLE 3.
FT DOMAIN 291 376 APPLE 4.
FT DOMAIN 389 621 CATALYTIC.
FT CARBOHYD 127 127 N-LINKED (GLCNAC. . .).
FT CARBOHYD 308 308 N-LINKED (GLCNAC. . .).
FT CARBOHYD 308 308 N-LINKED (GLCNAC. . .).
FT CARBOHYD 336 336 N-LINKED (GLCNAC. . .).
FT CARBOHYD 453 453 N-LINKED (GLCNAC. . .).
FT CARBOHYD 494 494 N-LINKED (GLCNAC. . .).
FT ACT_SITE 434 434 CHARGE RELAY SYSTEM.
FT ACT_SITE 483 483 CHARGE RELAY SYSTEM.
FT ACT_SITE 578 578 CHARGE RELAY SYSTEM.
FT ACT_SITE 77 77 CHARGE RELAY SYSTEM.
FT DISULFID 21 104
FT DISULFID 47 77
FT DISULFID 51 57
FT DISULFID 111 194
FT DISULFID 137 166
FT DISULFID 141 147
FT DISULFID 201 284
FT DISULFID 227 256
FT DISULFID 231 237
FT DISULFID 292 375
FT DISULFID 318 347
FT DISULFID 322 328
FT DISULFID 340 345
FT DISULFID 383 503
FT DISULFID 419 435
FT DISULFID 517 584
FT DISULFID 548 563
FT DISULFID 574 602
SQ SEQUENCE 638 AA; 71369 MW; E62F9C1053838FB4 CRC64;

Query Match 21.0%; Score 569.5; DB 1; Length 638;
Best Local Similarity 32.8%; Pred. No. 3.7e-35;
Matches 155; Conservative 64; Mismatches 160; Indels 93; Gaps 23;

OY 55 RVLQASNPVYCTQPKSPSGTQVCTSKTKALCTTLTGTFVLGALALAGLMLKFMGS--- 111
Db 218 RVLTL--PDAFVCR-----TICTYHPN--CLFPEFYR-----NWKIESQRNV 255
OY 112 ---KCSNGSGLTCDSS-----SG---TCINP-----SNMCDGVSHCGGDEENRCVRL 151
Db 256 CLKTKTSSTSPSSSTPQENTISGYSLLTKRRLPEPCHSKRIYGVDF--GGBELN----- 308
OY 152 YGDNFLLOYSSGQRKWHVPCODDMNENYGAACRDMGYKKNPYSSQGIVDSSGTS--- 208
Db 309 --VTYKGV-----NVQDE-----TCTM-IRCOFFYISLPEDCKEKCKC 347
OY 209 FKLNTSAGNVDIYKRLYHSDACSSKRAVYSLRCIACGVN--INSSRSHRIVGESALPGA 266

```



```

Db 348 FLRLMDGSPRTI---AVTQGGSSG---YSLRLCTMGDMSVCTKTSTRIYVGTNSMGE 401
Oy 267 WPMOVSILHVQ---NVHVGGSITTEBMVITAHCHYE-KRLNPNMHTAPAGILROSMFY 322
Db 402 WPMOVSLOVKRLTAQRHLLCGSSLIQHGWHTAAHCEDEGLPDQVWR--IYSGILNLSDTIK 459
Oy 323 GAGY-QVEKVISHPNVDSKTKNNNDALMLKQPLTFNDLVKRPVCLPFPNGMMLQPEOLCI 381
Db 460 DHPFQIOKEIILIIHQYKXKSEGNHDLALIKLQALPNTPEQRKICLPSKGDSTIYTNCAV 519
Oy 362 SGWATTEERKGTSEVLNAKAVLLIETORCNSRYVDNLITTPAMICAGFLQGVNDSQGDG 441
Db 520 TGMGFSKKEGELONTLOKRVNIPLVYNEEQKRY-ODKYITQMVYAGAGYKEGKDKDGDS 578
Oy 442 GGPLTTSKNNIMLIGDTSWGSCGAKAVPYGYGVNMYTDTIYROMR-ADG 492
Db 579 GGPLVCKRNGMRLVIGITSMGECARREQPGVYTYVAETMDILKETOSSDG 630

RESULT 9
HEPS_HUMAN
ID HEPS_HUMAN STANDARD: PRT: 417 AA.
AC P05981:
DT 01-NOV-1988 (Rel. 09, Created)
DT 01-NOV-1988 (Rel. 09, Last sequence update)
DT 01-NOV-1997 (Rel. 33, Last annotation update)
DE SERINE PROTEASE HEPsin (EC 3.4.21.-) (TRANSMEMBRANE PROTEASE, SERINE
DE 1).
GN HPN OR TMPRSS1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
OX NCBI_Taxid:9606;
RN [1]
RP SEQUENCE FROM N.A.
RP TISSUE=Liver:
RC MEDLINE=88209431; PubMed=2835076;
RX Leytus S.P., Loeb K.R., Hagen F.S., Kurachi K., Davie E.W.;
RA "A novel trypsin-like serine protease (hepsin) with a putative
RT transmembrane domain expressed by human liver and hepatoma cells."
RL Biochemistry 27:1067-1074(1988).
RN [2]
RP CHARACTERIZATION.
RX MEDLINE=91358502; PubMed=1885621;
RA Tsuji A., Torres-Rosado A., Arai T., le Beau M.M., Lemons R.S.,
RA Chou S.H., Kurachi K.;
RT "Hepsin, a cell membrane-associated protease. Characterization,
RT tissue distribution, and gene localization."
RL J. Biol. Chem. 266:16948-16953(1991).
RN [3]
RP CHARACTERIZATION.
RX MEDLINE=93348237; PubMed=834633;
RA Torres-Rosado A., O'Shea K.S., Tsuji A., Chou S.H., Kurachi K.;
RT "Hepsin, a putative cell-surface serine protease, is required for
RT mammalian cell growth."
RL Proc. Natl. Acad. Sci. U.S.A. 90:7181-7187(1993).
CC -1- FUNCTION: PLAYS AN ESSENTIAL ROLE IN CELL GROWTH AND MAINTENANCE
CC OF CELL MORPHOLOGY.
CC -1- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN.
CC -1- TISSUE SPECIFICITY: PRESENT IN MOST TISSUES, WITH THE HIGHEST
CC LEVEL IN LIVER.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1, ALSO KNOWN AS THE
CC TRYPsin FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC EMBL; M18930; AAA36013.1; -.

```

DR	EMBL; X07732; CAA30558.1; -.
DR	EMBL; X07002; CAA30058.1; -.
DR	PIR; S00845; S00845.
DR	HSSP; P00763; IDPO.
DR	MEROPS; S01.224; -.
DR	MIM; 142440; -.
DR	InterPro; IPR001254; -.
DR	InterPro; IPR001314; -.
DR	Pfam; PF00089; trypsin_1
DR	PRINTS; PR00722; CHYMOTRYPSIN.
DR	PROSITE; PS00134; TRYPSIN_HIS.1.
DR	PROSITE; PS00135; TRYPSIN_SER.1.
KW	Hydrolase; Serine protease; Transmembrane; Signal-anchor.
FT	CHAIN 1 162
FT	CHAIN 163 417
FT	DOMAIN 1 17
FT	TRANSMEM 18 44
FT	
FT	DOMAIN 45 417
FT	DOMAIN 163 417
FT	ACT_SITE 203 203
FT	ACT_SITE 257 257
FT	ACT_SITE 353 353
FT	DISULFID 153 277
FT	DISULFID 188 204
FT	DISULFID 322 338
FT	DISULFID 349 381
FT	CARBOHYD 112 112
SO	SEQUENCE 417 AA; 45011 MW; B208FF61E53D7 CRC64;
	N-LINKED (GLCNAC...) (POTENTIAL).
	EXTRACELLULAR (POTENTIAL).
	CATALYTIC.
	CHARGE RELAY SYSTEM (BY SIMILARITY).
	CHARGE RELAY SYSTEM (BY SIMILARITY).
	CHARGE RELAY SYSTEM (BY SIMILARITY).
	INTERCHAIN (BY SIMILARITY).
	BY SIMILARITY.
	BY SIMILARITY.
	BY SIMILARITY.

Query Match	20.8%;	Score 565.5;	DB 1;	Length 417;
Best Local Similarity	30.7%;	Pred. No. 4.6e-35;		
Matches 133;	Conservative 68;	Mismatches 173;	Indels 59;	Gaps 9

ID	ST14_HUMAN	STANDARD:	PRT:	855 AA.
QY	77	CTSKTKNALCITLLTGFLTVGAALAAAGLIMFKMGSCKSNSIEDSDSGTCINPSNMCDGV	138	
Db	12	CCSRKVA--ALTAGTLLTLTAIGA-----ASMAIVA	41	
QY	137	SHCGEGEDENRCVRLYGFNFLLQVYSSQSKSGMHVPCODDMNENYGRAACRDMGYKNFYS	199	
Db	42	VLRSDQEPFLPYGVSSADARLWYFDKTEGTWRLLCSSRAARVAGLSCDEMGLRALTH	101	
QY	197	SQGIYDDSGSTSPFKLNTSAGNVDTYK-KLYHSDA-----CSSKAVYSLEFCTACGV	248	
Db	102	SELVYRTAGA-----NQTSGFEFCVDBGRPLPHTQRLLEVISVCCDCCPGRGFLAACDCDG-	154	
QY	247	NLNSRSRGIYGESALPLGCAPMOVSLHYQVAVHCGGSIIPETIVYAAHCVEKFLANPW	308	
Db	155	-RRRLPVDRIYGVGADTSLGRMPMOVSLRYDCAHLCCGSSLLSGDVPVLTAAHCFFERNVLS	213	
QY	307	HWTAFAGILRQSEMFYGAGYVEKVISHPN-----DSKTKNDIALMKLQKPLETNDL	366	
Db	214	RMRFAGAVAAQASP-HGLQLGVQAVVYHGGYLPERDPENSENSDILLVHLSPLPLEY	272	
QY	361	VKPVCLPMPGMMQLPEQOLCMTISGATGEKEKSTSEVINAVALLLIETQRCSRYVYNLI	420	
Db	273	IQPVCLPAGALVDKCLCTYVTGNGNTQYVQQAGVLDIAEAVPLTISNDVCGADFTNQI	332	
QY	421	TPAIVICAGFLQGNVDSCGGSDGSPLV---TSKNNIMWLIGDTISVSGCAKAYRPGVYN	476	
Db	333	KPKMEFCAGYPEGGIDACGCGDGSPFVCEDSISRPRMRLGIVSMGICALQAPGVYTK	392	
QY	477	VWVFTDMYIYRQMR	489	
Db	393	VSDREFWIFQAIK	405	

DT 01-OCT-2000 (Rel. 40, Last sequence update)
 DT 01-OCT-2000 (Rel. 40, Last annotation update)
 DE SUPPRESSOR OF TUMORIGENICITY 14 (EC 3.4.21.-) (MATRIPTASE) (MEMBRANE-
 DE TYPE SERINE PROTEASE 1) (MT-SPI).
 GN ST14 OR PRS14 OR SRC19.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA MEDLINE-99303581; PubMed-10373424;
 RA Lin C.Y., Anders J., Johnson M., Sang Q.A., Dickson R.B.;
 RT "Molecular cloning of cDNA for matrilysin, a matrix-degrading serine
 RT protease with trypsin-like activity."
 RL J Biol. Chem. 274:18231-18236(1999).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Takeuchi T., Shuman M.A., Craik C.S.;
 RT "Reverse biochemistry: Use of macromolecular protease inhibitors to
 RT dissect complex biological processes and identify a membrane-type
 RT serine protease in epithelial cancer and normal tissue."
 RL Proc. Natl. Acad. Sci. U.S.A. 96:11054-11061(1999).
 RN [3]
 RP CHARACTERIZATION.
 RC TISSUE-Milk;
 RX PubMed-10373425;
 RA Lin C.Y., Anders J., Johnson M., Dickson R.B.;
 RT "Purification and characterization of a complex containing matrilysin
 RT and a Kunitz-type serine protease inhibitor from human milk."
 RL J. Biol. Chem. 274:18237-18242(1999).
 CC -1- FUNCTION: DEGRADERS EXTRACELLULAR MATRIX. PROPOSED TO PLAY A ROLE
 CC IN BREAST CANCER INVASION AND METASTASIS. EXHIBITS TRYPSIN-LIKE
 CC ACTIVITY AS DEFINED BY CLEAVAGE OF SYNTHETIC SUBSTRATES WITH ARG
 CC OR LYS AS THE P1 SITE.
 CC -1- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN (PROBABLE).
 CC -1- SIMILARITY: CONTAINS 4 LDL-RECEPTOR CLASS A DOMAINS.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
 CC TRYPSIN FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/)
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL: AF118224; AAD42765.2; -;
 DR EMBL: AF133086; AAF00109.1; -;
 DR HSSP: P00763; IDPO.
 DR MEROPS: S01.302; -;
 DR InterPro: IPR000859; -;
 DR InterPro: IPR001254; -;
 DR InterPro: IPR001314; -;
 DR InterPro: IPR002172; -;
 DR Pfam: PF00057; 1d1_recept_a; 4.
 DR Pfam: PF00089; trypsin; 1.
 DR Pfam: PF00431; CUB; 2.
 DR PRINTS: PR00261; LDLRECEPTOR.
 DR PRINTS: PR00722; CHYMOTRYPSIN.
 DR PROSITE: PS00134; TRYPSIN_HIS; 1.
 DR PROSITE: PS00135; TRYPSIN_SER; 1.
 DR PROSITE: PS01180; CUB; 2.
 DR PROSITE: PS01209; LDLRA_1; 2.
 DR PROSITE: PS00068; LDLRA_2; 4.
 DR Signal.anchor: Glycoprotein; Hydrolase; Serine protease;
 KW Transmembrane; Repeat.
 FT DOMAIN 1 55 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 56 76 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
 FT (POTENTIAL).
 FT DOMAIN 77 855 EXTRACELLULAR (POTENTIAL).

FT DOMAIN 214 334 CUB 1.
 FT DOMAIN 340 447 CUB 2.
 FT DOMAIN 452 487 LDL-RECEPTOR CLASS A 1.
 FT DOMAIN 487 524 LDL-RECEPTOR CLASS A 2.
 FT DOMAIN 524 560 LDL-RECEPTOR CLASS A 3.
 FT DOMAIN 566 603 LDL-RECEPTOR CLASS A 4.
 FT DOMAIN 614 851 CATALYTIC.
 FT ACT_SITE 656 656 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 711 711 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 805 805 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT CARBOHYD 109 109 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 302 302 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 485 485 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 772 772 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 855 AA; 94769 MW; 26143132C01F99C9 CRC64;
 Query Match 20.6%; Score 558.5; DB 1; Length 855;
 Best Local Similarity 33.9%; Pred. No. 3.4e-34;
 Matches 131; Conservative 57; Mismatches 129; Indels 69; Gaps 14;
 QY 110 GSKCSNSGIECDSSGTCINPSMNCDCVSHCPGGEDENRKCYRLGPNFTILOVYSQKSMH 169
 DB 522 GCSCPAQTERFC-SNGKCLSKSQCNCGKDCGDSDEASCPKV---NVV----- 565
 QY 170 PVOQDDMNENYGRACRDMGKKNFYSSOGIYDSDGSTSPKNTSAGVNDYKKLYHSD 229
 DB 566 -----TCT-----KHTYRLNG-----LCLSKGNECDKDCSD 595
 QY 230 ACSKRAVSLRACIACVNLNS-SRQRIYGESALPGAMPQVSLH-VONVHCGSGIIT 287
 DB 596 GSDEK-----DCDGLRFSFTRQARVVGTDADGEMPMQVSLHALGCHIGCASLIS 647
 QY 288 PEVITVAARCVKPEP-----LNNPMTAFAGILRQSMFAGAQY---VEKVISHPYDSK 340
 DB 648 PMLVSAARCYIDDRGFRYSDPQWTAFLGLHDQS-QRSAPGQERRLRRIISHPEFNEF 706
 QY 341 TKKNDLAIKLOKLEFNFIDKPVCLPNDGMLQPOLCISQMGATEKKGSEVLNA 400
 DB 707 TPDYDIALLELEKPAEYSSMVRPCLPDASHVPACKAIWYTMGHTQYGGICALILQNG 766
 QY 401 KVLIEFORNSRYVDNLTTPAMICAGFLOGNVDSQGDGSGPL-VYSKNIMWLIGD 458
 DB 767 EIRVINGTTCEN-LLPQGITPRMCGVFLSGVDSQGDGSGPLSSVEADGRI-FQAGV 823
 QY 459 TSWGSCAKAYRPGVYGNVWFTDWI 484
 DB 824 VSWGDCCAQRNKPQVYTRLPFRDWI 849
 RESULT 11
 KAL_RAT STANDARD; PRT; 638 AA.
 AC P14272;
 DT 01-JAN-1990 (Rel. 13, Created)
 DT 01-JAN-1990 (Rel. 13, Last sequence update)
 DT 01-OCT-2000 (Rel. 40, Last annotation update)
 DE PLASMA KALLIKREIN PRECURSOR (EC 3.4.21.34) (PLASMA PREKALLIKREIN)
 DE (KININOGENIN) (FLETCHER FACTOR).
 GN Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA MEDLINE-91129236; PubMed-1993180;
 RA Beaudien G., Rosinski-Chupin I., Mattei M.-G., Mbikay M., Chretien M.,
 RA Seidah N.G.;
 RT "Gene structure and chromosomal localization of plasma kallikrein."
 RL Biochemistry 30:1628-1635(1991).
 RN [2]
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.

RA MEDLINE=90091743; PubMed=2598771;
RA Seidah N.G., Ladenheim R., Mbikay M., Hamelin J., Lutfalla G.,
RA Rougeon F., Lazou C., Chretien M.;
RT "The cDNA structure of rat plasma kallikrein";
RL DNA 8:563-574(1989).
CC -1- FUNCTION: THE ENZYME CLEAVES LYS-ARG AND ARG-SER BONDS. IT
CC ACTIVATES, IN A RECIPROCAL REACTION, FACTOR XII AFTER ITS BINDING
CC TO A NEGATIVELY CHARGED SURFACE. IT ALSO RELEASES BRADYKININ FROM
CC HWM KININOGEN AND MAY ALSO PLAY A ROLE IN THE RENIN-ANGIOTENSIN
CC SYSTEM BY CONVERTING PRORENIN INTO RENIN.
CC -1- SUBUNIT: THE ZYMOGEN IS ACTIVATED BY FACTOR XIIa, WHICH CLEAVES
CC THE MOLECULE INTO A LIGHT CHAIN, WHICH CONTAINS THE ACTIVE SITE,
CC AND A HEAVY CHAIN, WHICH ASSOCIATES WITH HWM KININOGEN. THESE
CC CHAINS ARE LINKED BY ONE OR MORE DISULFIDE BONDS.
CC -1- SIMILARITY: CONTAINS 4 APPLE DOMAINS.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC TRYPSIN FAMILY. PLASMA KALLIKREIN SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M62357; AAA74563.1; -;
DR EMBL; M62358; AAA74563.1; JOINED.
DR EMBL; M62346; AAA74563.1; JOINED.
DR EMBL; M62347; AAA74563.1; JOINED.
DR EMBL; M62349; AAA74563.1; JOINED.
DR EMBL; M62350; AAA74563.1; JOINED.
DR EMBL; M62351; AAA74563.1; JOINED.
DR EMBL; M62352; AAA74563.1; JOINED.
DR EMBL; M62353; AAA74563.1; JOINED.
DR EMBL; M62354; AAA74563.1; JOINED.
DR EMBL; M62355; AAA74563.1; JOINED.
DR EMBL; M62356; AAA74563.1; JOINED.
DR EMBL; M30282; AAA41463.1; -;
DR EMBL; M58590; AAA42069.1; -;
DR PIR; A39180; KORTPL.
DR HSSP; P00750; LRFE.
DR MEROPS; S01.212; -;
DR InterPro; IPR000177; -;
DR InterPro; IPR001254; -;
DR InterPro; IPR001314; -;
DR InterPro; IPR003014; -;
DR Pfam; PF00024; PAN; 4.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PRO0005; APPLEDOMAIN.
DR PRINTS; PRO0722; CHYMOTRYPSIN.
DR PROSITE; PS00134; TRYPSIN_HIS; 1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
DR PROSITE; PS00495; APPLE; 4.
KW Hydrolyase; Serine protease; Glycoprotein; Plasma; zymogen; Signal;
KW Fibrinolysis; Blood coagulation; Inflammatory response; Liver;
KW Repeat.
FT SIGNAL 1 19
FT CHAIN 20 390 PLASMA KALLIKREIN HEAVY CHAIN.
FT CHAIN 391 638 PLASMA KALLIKREIN LIGHT CHAIN.
FT DOMAIN 20 105 APPLE 1.
FT DOMAIN 110 195 APPLE 2.
FT DOMAIN 200 285 APPLE 3.
FT DOMAIN 291 376 APPLE 4.
FT DOMAIN 389 621 CATALYTIC.
FT CARBOHYD 127 127 N-LINKED (GLCNAC. . .) (PROBABLE).
FT CARBOHYD 308 308 N-LINKED (GLCNAC. . .) (PROBABLE).
FT CARBOHYD 396 396 N-LINKED (GLCNAC. . .) (PROBABLE).
FT CARBOHYD 453 453 N-LINKED (GLCNAC. . .) (PROBABLE).
FT CARBOHYD 494 494 N-LINKED (GLCNAC. . .) (PROBABLE).
FT ACT_SITE 434 434 CHARGE RELAY SYSTEM.
FT ACT_SITE 483 483 CHARGE RELAY SYSTEM.
FT ACT_SITE 578 578 CHARGE RELAY SYSTEM.

FT DISULFD 21 104 BY SIMILARITY.
FT DISULFD 47 77 BY SIMILARITY.
FT DISULFD 51 57 BY SIMILARITY.
FT DISULFD 111 194 BY SIMILARITY.
FT DISULFD 137 166 BY SIMILARITY.
FT DISULFD 141 147 BY SIMILARITY.
FT DISULFD 201 284 BY SIMILARITY.
FT DISULFD 227 256 BY SIMILARITY.
FT DISULFD 231 237 BY SIMILARITY.
FT DISULFD 292 375 BY SIMILARITY.
FT DISULFD 318 347 BY SIMILARITY.
FT DISULFD 322 328 BY SIMILARITY.
FT DISULFD 340 345 BY SIMILARITY.
FT DISULFD 383 503 BY SIMILARITY.
FT DISULFD 419 435 BY SIMILARITY.
FT DISULFD 517 584 BY SIMILARITY.
FT DISULFD 548 563 BY SIMILARITY.
FT DISULFD 574 602 BY SIMILARITY.
SO SEQUENCE 638 AA; 71273 MW; 454BB27E8CA8F98 CRC64;
Query Match 20.5%; Score 556; DB 1; Length 638;
Best local similarity 36.6%; Pred. No. 3.8e-34;
Matches 117; Conservative 59; Mismatches 128; Indels 16; Gaps 7;
QY 181 GRACRDMGKYN--NFYSOGIYVD--SGSTFMKNTSAGNVDYKLYHSDACSK 234
DB 314 GADACQETCTRTIRCQFTYSLPDDCKAECKCSLRSTDSGSPRTIYEAGSSGYSLR 373
QY 235 AVSLRCIACGVNLSNRSQSRIVGESALPGAMPQVSLHQN--HVCGSITTEPMI 291
DB 374 LCKYESSDCCTKIN---ARIVGNTSSILEMPQVSLQKLVQSQNMCGSIIGRWI 429
QY 292 VTAACHVEKPLNNPWHMTAFAGILROSEFMFGAGY-QVEKVISHPNVDYKRNNDIALMK 350
DB 430 LTAHCFD-GIPYDPVMTIYGILNLSITKTPTSSKELIHKYKMSGSDIALIK 488
QY 351 LQKPTENDLVKPYCLPMPGMLOPEQLCWTISGWGATEEKKTSEVLNAKVLLETRC 410
DB 489 LQPLNTYEFQKPICLPSCADPTNTYTNCWVTGWTGERETONILQKATIPVPMEEC 548
QY 411 NSRYVDNLITPMACAEFLQGNVDSGSGPVTSKNNIMWLDIGTSGSCAKAYR 470
DB 549 OKY-RDVIYITKMCACYKREGIDACKGSGPLVCKHSGRWLVITSGEGCAREQ 607
QY 471 PGVYGNAVFTDWIYQKRA 490
DB 608 PGVITKVAEYIDWILEKIQS 627
RESULT 12
ID ST14_MOUSE STANDARD; PRT; 902 AA.
AC P56677;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE SUPPRESSOR OF TUMORIGENICITY 14 (EC 3.4.21.-) (EPTIHIN).
OS ST14 OR PSS14.
GN Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxId=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C.B.17SCID; TISSUE=Thymus;
RA PubMed=10199918;
RA Kim M.G., Chen C., Lyu M.S., Cho E.G., Park D., Kozak C.,
RA Schwartz R.H.;
RT *Cloning and chromosomal mapping of a gene isolated from thymic
RT stromal cells encoding a new mouse type II membrane serine protease,
RT eptih, containing four LDL receptor modules and two CUB domains";
RL Immunogenetics 49:420-428(1999).

```

CC -1- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN (PROBABLE).
CC -1- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN INTESTINE, KIDNEY, LUNG,
CC AND THYMUS. NOT EXPRESSED IN SKELETAL MUSCLE, LIVER, HEART,
CC TESTIS, AND BRAIN.
CC -1- SIMILARITY: CONTAINS 4 LDL-RECEPTOR CLASS A DOMAINS.
CC -1- SIMILARITY: CONTAINS 2 CUB DOMAINS.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1, ALSO KNOWN AS THE
CC TRYPSIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AF042822; AAD02230.1; -.
CC MGD: MGI:1338881; Scl14.
CC MEROPS: S01.014; -.
CC InterPro: IPR000859; -.
CC InterPro: IPR001254; -.
CC InterPro: IPR001314; -.
CC InterPro: IPR002172; -.
CC Pfam: PF00431; CUB; 2.
CC Pfam: PF00057; ldl_recept_a; 4.
CC Pfam: PF00089; trypsin_1.
CC PRINTS: PR00261; LDLRECEPTOR.
CC PRINTS: PR00722; CHYMOTRYPSIN.
CC PROSITE: PS01209; LDLRA_1; 2.
CC PROSITE: PS50068; LDLRA_2; 4.
CC PROSITE: PS01180; CUB; 2.
CC PROSITE: PS00134; TRYPSIN_HIS; 1.
CC PROSITE: PS00135; TRYPSIN_SER; 1.
CC Signal-anchor: Glycoprotein; Hydrolase; Serine protease;
CC Transmembrane; Repeat.
CC TRANSMEM 1 55
CC DOMAIN 56 76
CC FT 77 902 CYTOPLASMIC (POTENTIAL).
CC FT 331 331 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
CC FT 340 444 (POTENTIAL).
CC FT 451 488 EXTRACELLULAR (POTENTIAL).
CC FT 489 522 CUB 1.
CC FT 523 561 LDL-RECEPTOR CLASS A 1.
CC FT 565 604 LDL-RECEPTOR CLASS A 2.
CC FT 614 851 LDL-RECEPTOR CLASS A 3.
CC FT 851 851 CATALYTIC.
CC FT ACT_SITE 656 656 CHARGE RELAY SYSTEM (BY SIMILARITY).
CC FT ACT_SITE 711 711 CHARGE RELAY SYSTEM (BY SIMILARITY).
CC FT ACT_SITE 805 805 CHARGE RELAY SYSTEM (BY SIMILARITY).
CC FT CARBOHD 107 107 N-LINKED (GLCNAC...) (POTENTIAL).
CC FT CARBOHD 302 302 N-LINKED (GLCNAC...) (POTENTIAL).
CC FT CARBOHD 365 365 N-LINKED (GLCNAC...) (POTENTIAL).
CC FT CARBOHD 421 421 N-LINKED (GLCNAC...) (POTENTIAL).
CC FT CARBOHD 489 489 N-LINKED (GLCNAC...) (POTENTIAL).
CC FT CARBOHD 772 772 N-LINKED (GLCNAC...) (POTENTIAL).
CC SEQUENCE 902 AA; 99645 MW; 2A3ED02AF69C04E CR664;

```

```

Query Match 20.2% Score 548.5; DB 1; Length 902;
Best Local Similarity 32.7%; Pred. No. 2e-33;
Matches 123; Conservative 66; Mismatches 118; Indels 69; Gaps 13;

```

```

QY 110 GSKGNSGTECHSSSTGCTIPMSWMCQVSHCPGEGEDENRCVRLXGPFILQVYSSQKRSWH 169
DB 522 GCSGAGSEFC-SNKGCLLPQSQKCNKGCGSGSCASC----- 559
QY 170 PVCQDDMNTNRYACRACRMDGYK--NNFYSSGIVDDSGSTSPMKLNTSAGNVDIYKLKH 227
DB 560 -----DSNVVSGCTKTYTKCQNGCLCSKNGPECDGTD----- 593
QY 228 SDAGSSKAVVSLRCLACGVNLSSRSRIVGESALPGAMPQVSLH-VQNVHVCGSIT 286

```

```

DB 594 SDGSDK-----NC-DCGLR-SFTKQARVVGTTNADEGEMPMQVSLHALGCHLGSASLI 646
QY 287 TPETWYTAHCHVEKPLANNWH-----WTFAGILRQSFMYGAGYQ---VEKYSHPNYS 339
DB 647 SPDWLSAAHCFQDDKNRKYSDYTMATFLGLDQD-KRSASGVQKLKRLITTHSFND 705
QY 340 KTKNDIALMLOKPLTFEDLVKPYCLPMPMDLOPEQOLCWSMGATEKGTSEVLNA 399
DB 706 FTFVDIALLELEKSYETVVRPILCPDPAHFVPRGKAIWGTNGHTEGCGALILQK 765
QY 400 AKVLIETRCNSRYVDNLTPPMICAGFLOGNDSCGDSGGLVYS-KNNIMWLID 458
DB 766 GEIRVINGTCED--LMPQIIPRMVCVGLSGVDSGCGDGPLSAEKDGRMFGVY 823
QY 459 TSWGSCAKAYRPGVY 474
DB 824 VSWGEGCAQRNRPVGY 839

RESULT 13
KAL_MOUSE
ID KAL_MOUSE STANDARD; PRT; 638 AA.
AC P26262;
DT 01-MAY-1992 (Rel. 22, Created)
DT 01-MAY-1992 (Rel. 22, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE PLASMA KALLIKEIN PRECURSOR (EC 3.4.21.34) (PLASMA PREKALLIKEIN)
DE (KININOGENIN) (FLETCHER FACTOR).
GN KUK3 OR PK.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_Taxid=10090;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC STRAIN-BALB/C; TISSUE=Liver;
RX MEDLINE=91090844; PubMed=2264928;
RA Seidah N.G., Sawyer N., Hamelin J., Mbikay M., Chretien M.,
RA Brachpapa L., Rochemont J., Mbikay M., Chretien M.,
RT "Mouse plasma kallikrein: cDNA structure, enzyme characterization,
RT and comparison of protein and mRNA levels among species."
RL DNA Cell Biol. 9:737-748(1990).
CC -1- FUNCTION: THE ENZYME CLEAVES LYS-ARG AND ARG-SER BONDS. IT
CC TO A NEGATIVELY CHARGED SURFACE. IT ALSO RELEASES BRADYKININ FROM
CC HMW KININOGEN AND MAY ALSO PLAY A ROLE IN THE RENIN-ANGIOTENSIN
CC SYSTEM BY CONVERTING PRORENIN INTO RENIN.
CC -1- SUBUNIT: THE ZYMOGEN IS ACTIVATED BY FACTOR XIIA, WHICH CLEAVES
CC THE MOLECULE INTO A LIGHT CHAIN, WHICH CONTAINS THE ACTIVE SITE,
CC AND A HEAVY CHAIN, WHICH ASSOCIATES WITH HMW KININOGEN. THESE
CC CHAINS ARE LINKED BY ONE OR MORE DISULFIDE BONDS.
CC -1- SIMILARITY: CONTAINS 4 APPLE DOMAINS.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1, ALSO KNOWN AS THE
CC TRYPSIN FAMILY. PLASMA KALLIKEIN SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: M58588; AAA63393.1; -.
CC PIR: A36557; KQMSPL.
CC HSSP: P00750; 1RTP.
CC MEROPS: S01.212; -.
CC MGD: MGI:102849; Klk3.
CC InterPro: IPR00177; -.
CC InterPro: IPR001254; -.
CC InterPro: IPR001314; -.
CC InterPro: IPR003014; -.
CC Pfam: PF00024; PAN; 4.

```

DR Pfam; PF00089; trypsin; 1.
 DR PRINTS; PR00005; APPELDOMAIN.
 DR PRINTS; PR00722; CHYMOTRYPSIN.
 DR PROSITE; PS00134; TRYPSIN_HIS; 1.
 DR PROSITE; PS00135; TRYPSIN_SER; 1.
 DR PROSITE; PS00495; APPEL; 4.
 DR Hydrolyase; Serine protease; Glycoprotein; Plasma; Zymogen; Signal;
 KW Fibrinolysis; Blood coagulation; Inflammatory response; Liver;
 KM Repeat.
 FT SIGNAL 1 19
 FT CHAIN 20 390 PLASMA KALLIKREIN HEAVY CHAIN.
 FT DOMAIN 391 638 PLASMA KALLIKREIN LIGHT CHAIN.
 FT DOMAIN 20 105 APPLE 1.
 FT DOMAIN 110 195 APPLE 2.
 FT DOMAIN 200 285 APPLE 3.
 FT DOMAIN 291 376 APPLE 4.
 FT DOMAIN 389 621 CATALYTIC.
 FT CARBOHYD 127 127 N-LINKED (GLCNAC. . .) (PROBABLE).
 FT CARBOHYD 308 308 N-LINKED (GLCNAC. . .) (PROBABLE).
 FT CARBOHYD 396 396 N-LINKED (GLCNAC. . .) (PROBABLE).
 FT CARBOHYD 453 453 PROBABLE.
 FT CARBOHYD 494 494 N-LINKED (GLCNAC. . .) (PROBABLE).
 FT ACT_SITE 434 434 CHARGE RELAY SYSTEM.
 FT ACT_SITE 483 483 CHARGE RELAY SYSTEM.
 FT ACT_SITE 578 578 CHARGE RELAY SYSTEM.
 FT ACT_SITE 21 104 CHARGE RELAY SYSTEM.
 FT DISULFID 47 77 BY SIMILARITY.
 FT DISULFID 51 57 BY SIMILARITY.
 FT DISULFID 111 194 BY SIMILARITY.
 FT DISULFID 137 166 BY SIMILARITY.
 FT DISULFID 141 147 BY SIMILARITY.
 FT DISULFID 201 284 BY SIMILARITY.
 FT DISULFID 227 256 BY SIMILARITY.
 FT DISULFID 231 237 BY SIMILARITY.
 FT DISULFID 292 375 BY SIMILARITY.
 FT DISULFID 318 347 BY SIMILARITY.
 FT DISULFID 322 328 BY SIMILARITY.
 FT DISULFID 340 345 BY SIMILARITY.
 FT DISULFID 383 503 BY SIMILARITY.
 FT DISULFID 419 435 BY SIMILARITY.
 FT DISULFID 517 584 BY SIMILARITY.
 FT DISULFID 548 563 BY SIMILARITY.
 FT DISULFID 574 602 BY SIMILARITY.
 FT SEQUENCE 638 AA; 71368 MW; CC27C93AB1086599 CRC64;

Query Match 19.8%; Score 538; DB 1; Length 638;
 Best Local Similarity 37.1%; Pred. No. 8.3e-33;
 Matches 115; Conservative 52; Mismatches 117; Indels 26; Gaps 6;
 185 CRDMGKNNFYSSOGIVDDSGSTSFMLKNTSAGNVDYKLYHSDACSSKAVALSLRCLAC 244
 340 CKERGGKCSLRLS---FDGSPTRITYGQSSGSLRCLKLVDSPTDKT----- 387
 245 GVNINSSQSHIVGESLPGAMPVQVSLYQV---HVGCGSITTPWTYTAHCVKRP 301
 388 -----NATIVGCTNASTLSGEMPMOVSLOYKLVSQLCGSIIIGROWVLTAAHCFD-G 438
 302 LNPDMHTAFAGILRSOFMEYGA-GYQVEKVISHPNDSKTKNDIDALMLKLOKPLENDL 360
 439 IPRYDVRIRIYGIIISLSEITETKPSRKELIHOEYKVSAGNDIALIKIQTPLANTYEF 498
 361 VKPVCLPNCMLQPEOLCWTISGATGEKCTSEVLNAKVLLETORNSRYVDNLI 420
 499 QKPICLPSKADNTIYNCVWTGMYTKREGGETONILQKATIPLPNECCQKTY-RDVI 557
 421 TPAMICAGFLDGNVDSGCGSGPLVTSKNITMILGDTSGSGCAKAYRPGVGNWVF 480
 558 NKOMICAGYKCGTADCKGDSGGLVCKHSGRMQLVGTISWGEGCGRDKDPGYTVKSEX 617
 481 TDVIYRQMA 490
 618 MDVILEKTQS 627

RESULT 14
 HEP5_MOUSE
 ID HEP5_MOUSE STANDARD; PRT: 416 AA.
 AC O35433;
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 15-JUL-1998 (Rel. 36, Last annotation update)
 DE SERINE PROTEASE HEP5IN (EC 3.4.21.-).
 GN HPN.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_Taxid=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Liver.
 RX MEDLINE=98058912; PubMed=9395459;
 RA Vu T.-K.H., Liu R.W., Haakma C., Tomasek J.J., Howard E.W.;
 RT "Identification and cloning of the membrane-associated serine
 protease, hepsin, from mouse preimplantation embryos."
 RL J. Biol. Chem. 272:31315-31320(1997).
 CC -1- FUNCTION: PLAYS AN ESSENTIAL ROLE IN CELL GROWTH AND MAINTENANCE
 CC OF CELL MORPHOLOGY.
 CC -1- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
 CC TRYPSIN FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; AF030065; AAB84221.1; -.
 DR HSSP; P00763; IDPO.
 DR MEROPS; S01.224; -.
 DR MGD; MGI:1196620; Hpn.
 DR InterPro; IPR001254; -.
 DR InterPro; IPR001314; -.
 DR Pfam; PF00089; trypsin; 1.
 DR PRINTS; PR00722; CHYMOTRYPSIN.
 DR PROSITE; PS00134; TRYPSIN_HIS; 1.
 DR PROSITE; PS00135; TRYPSIN_SER; 1.
 KW Hydrolyase; Serine protease; Transmembrane; Signal-anchor.
 FT CHAIN 1 161 NON-CATALYTIC CHAIN (POTENTIAL).
 FT DOMAIN 162 416 CATALYTIC CHAIN (POTENTIAL).
 FT TRANSMEM 1 16 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 17 43 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
 FT (POTENTIAL).
 FT DOMAIN 44 416 EXTRACELLULAR (POTENTIAL).
 FT DOMAIN 162 416 CATALYTIC.
 FT ACT_SITE 202 202 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 256 256 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 352 352 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT DISULFID 152 276 INTERCHAIN (BY SIMILARITY).
 FT DISULFID 187 203 BY SIMILARITY.
 FT DISULFID 321 337 BY SIMILARITY.
 FT DISULFID 348 380 BY SIMILARITY.
 FT CARBOHYD 111 111 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT SEQUENCE 416 AA; 44739 MW; 432194FF4004F848 CRC64;

Query Match 19.7%; Score 536.5; DB 1; Length 416;
 Best Local Similarity 30.6%; Pred. No. 6.7e-33;
 Matches 132; Conservative 63; Mismatches 180; Indels 57; Gaps 9;
 77 CTSTKRALCTTLTFLVGAALAAGLMKKMGKSGIEDSGGTINPSMWDGV 136
 11 CSRPRKVAALIVGTL-LFLTG-----IGAASMAIVT 40

```

OY 137 SHCPGDEDCRCVRLYGPFFILQVYSSQRKSWHPYQODDMMENYGRACRDMGYNKNEYS 196
DB 41 ILLQSDQDEFLYVQSLSPDQSRILAVLDKTBGTRRLCQSSNARNVAGLCEEGFLRALAH 100
OY 197 SOGIYDDSGSTSMKLNLSA-----GNVDYIKLYHSDA---CSSKAVYSLRCIACGVN 247
DB 101 SELDVRTAGAN-----GTSGFVCEDEGGLPLAQRLDLVSYDCDPRGRLLTATCDQCG-- 153
OY 248 LNSSQSRIVGCESSALPGAMPQVSLHQNVAHVCGSITTPPMIYTAHCVKPLNNPWH 307
DB 154 RRKLPEYRIVGQSDSLGRMPQVSLRDGTHLCGSLISGVMVLTAAHCFPERNRVLSR 213
OY 308 WFAFAGILRQSPMFYAGAGYVEKYSHPNY-----DSKTKNNIDALMLQKPLPFENDLY 361
DB 214 WVEFGAVARISP-HAVVLGVDAVIYHGGYLPFRDPPTIDENSNDIALVHLSLPLETYI 272
OY 362 KPVCLPMPQMLQAPOLCWSGMGATEEKTSSEVLNAKAVLLIETORCNSRYVYDNLIT 421
DB 273 QVVCPLPAGQALVVDGKVCVTGMNMGTFYGGQAMVLQEARVPLISNEVCNSPDFYGNQIK 332
OY 422 PMITAGFLQGVNDSCQDGGGPLY-----TSKNNIWMILIGTSMGSCGAKAVRPGYGVN 477
DB 333 PMFAGYPEGGIDACQDGGSGPFCVCEDSISGTSRMLCGIIVSWGTCALARKPGYTVKV 392
OY 478 WFTDMYRMR 489
DB 393 TDFREMIKAIK 404

RESULT 15
FAIL_HUMAN
ID FAIL_HUMAN STANDARD: PRT: 625 AA.
AC P03951.
DT 23-OCT-1986 (Rel. 02, Created)
DT 23-OCT-1986 (Rel. 02, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE COAGULATION FACTOR XI PRECURSOR (EC 3.4.21.27) (PLASMA THROMBOPLASTIN
DE ANPECEDENT) (PTA).
CN F11.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP MEDLINE FROM N.A.
RX MEDLINE-86243360; PubMed=3636155;
RA Fujikawa K., Chung D.W., Hendrickson L.E., Davie E.W.;
RT "Amino acid sequence of human factor XI, a blood coagulation factor
RT with four tandem repeats that are highly homologous with plasma
RT prekallikrein."
RL Biochemistry 25:2417-2424(1986).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE-88107663; PubMed=2827746;
RA Asakai R., Davie E.W., Chung D.W.;
RT "Organization of the gene for human factor XI."
RN Biochemistry 26:7221-7228(1987).
RN [3]
RP PARTIAL SEQUENCE, AND DISULFIDE BONDS.
RX MEDLINE-91152017; PubMed=1998667;
RA McMullen B.A., Fujikawa K., Davie E.W.;
RT "Location of the disulfide bonds in human coagulation factor XI: the
RT presence of tandem apple domains."
RL Biochemistry 30:2056-2060(1991).
RN [4]
RP VARIANT LEU-301.
RX MEDLINE-90046656; PubMed=2813350;
RA Asakai R., Chung D.W., Ratnoff O.D., Davie E.W.;
RT "Factor XI (plasma thromboplastin antecedent) deficiency in Ashkenazi
RT Jews is a bleeding disorder that can result from three types of point
RT mutations."
RL Proc. Natl. Acad. Sci. U.S.A. 86:7667-7671(1989).

```

```

RN [5]
RP VARIANT LEU-301.
RX MEDLINE-92190478; PubMed=1547342;
RA Meijers J.C., Davie E.W., Chung D.W.;
RT "Expression of human blood coagulation factor XI: characterization of
RT the defect in factor XI type III deficiency."
RL Blood 79:1435-1440(1992).
CC -1- FUNCTION: FACTOR XI TRIGGERS THE MIDDLE PHASE OF THE INTRINSIC
CC PATHWAY OF BLOOD COAGULATION BY ACTIVATING FACTOR IX.
CC -1- CATALYTIC ACTIVITY: SELECTIVE CLEAVAGE OF ARG-1-ALA AND ARG-1-VAL
CC BONDS IN FACTOR IX TO FORM FACTOR IXA.
CC -1- SUBUNIT: HOMODIMER; LINKED BY A DISULFIDE BOND. AFTER ACTIVATION
CC THE HEAVY AND LIGHT CHAINS ARE ALSO LINKED BY A DISULFIDE BOND.
CC -1- PTM: ACTIVATED BY FACTOR XIIA (OR XII), WHICH CLEAVES EACH
CC POLYPEPTIDE AFTER ARG-387 INTO THE LIGHT CHAIN, WHICH CONTAINS THE
CC ACTIVE SITE, AND THE HEAVY CHAIN, WHICH ASSOCIATES WITH HIGH
CC MOLECULAR WEIGHT (HMW) KININOGEN.
CC -1- DISEASE: DEFECTS IN F11 ARE A CAUSE OF A BLOOD COAGULATION
CC ABNORMALITY (ROSENTHAL SYNDROME) OCCURRING IN HIGH FREQUENCY IN
CC ASHKENAZI JEWS.
CC -1- SIMILARITY: CONTRAINS 4 APPLE DOMAINS.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC TRYPSIN FAMILY. PLASMA KALLIKREIN SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M13142; AAA52487.1; -
DR EMBL: M20218; AAA51985.1; -
DR EMBL: M18296; AAA51985.1; JOINED.
DR EMBL: M21184; AAA51985.1; JOINED.
DR EMBL: M18298; AAA51985.1; JOINED.
DR EMBL: M18300; AAA51985.1; JOINED.
DR EMBL: M18301; AAA51985.1; JOINED.
DR EMBL: M18302; AAA51985.1; JOINED.
DR EMBL: M18303; AAA51985.1; JOINED.
DR EMBL: M18304; AAA51985.1; JOINED.
DR EMBL: M19417; AAA51985.1; JOINED.
DR EMBL: M20217; AAA51985.1; JOINED.
DR PIR: A27431; KFH01.
DR HSSP: P00763; IDPO.
DR MEROPS: S01.213; -.
DR MIM: 134540; -.
DR MIM: 264900; -.
DR InterPro: IPR000177; -.
DR InterPro: IPR001254; -.
DR InterPro: IPR001314; -.
DR InterPro: IPR003014; -.
DR Pfam: PF000024; PAN; 4.
DR Pfam: PF00089; trypsin; 1.
DR PRINTS: PR00005; APPLDOMAIN.
DR PRINTS: PR00722; CHYMOTRYPSIN.
DR PROSITE: PS00134; TRYPSIN_HIS; 1.
DR PROSITE: PS00135; TRYPSIN_SER; 1.
DR PROSITE: PS00495; APPL; 4.
KW Hydrolase; Serine protease; Glycoprotein; Plasma; Blood coagulation;
KW Duplication; Signal; Disease mutation.
FT SIGNAL 1 18
FT CHAIN 19 387
FT CHAIN 388 625
FT DOMAIN 19 104
FT DOMAIN 109 194
FT DOMAIN 199 284
FT DOMAIN 230 375
FT DOMAIN 384 625
FT DOMAIN 90 90
FT CARBOHYD 126 126
FT CARBOHYD 126 126

```

```

FT CARBOHYD 353 353 N-LINKED (GLCNAC. . .)
FT CARBOHYD 450 450 N-LINKED (GLCNAC. . .)
FT CARBOHYD 491 491 N-LINKED (GLCNAC. . .)
FT ACT_SITE 431 431
FT ACT_SITE 480 480
FT ACT_SITE 575 575
FT ACT_SITE 575 575
FT DISULFID 20 103
FT DISULFID 29 29 WITH A CYS RESIDUE.
FT DISULFID 46 76
FT DISULFID 50 56
FT DISULFID 110 193
FT DISULFID 136 165
FT DISULFID 140 146
FT DISULFID 200 283
FT DISULFID 226 255
FT DISULFID 230 236
FT DISULFID 291 374
FT DISULFID 317 346
FT DISULFID 321 327
FT DISULFID 339 339
FT DISULFID 380 500 INTERCHAIN.
                                INTERCHAIN (BETWEEN HEAVY AND LIGHT
                                CHAINS).
FT DISULFID 416 432
FT DISULFID 514 581
FT DISULFID 545 560
FT DISULFID 571 599
FT VARIANT 301 301
FT CONFLICT 226 226 F -> L (IN F11 DEFICIENCY).
                                /FTID-VAR_006622.
                                C -> S (IN REF. 2).
SQ SEQUENCE 625 AA; 70109 MM; 147AF94B709EBF CRC64;

```

Query Match Best Local Similarity 19.6%; Score 533; DB 1; Length 625;
Matches 122; Conservative 63; Mismatches 129; Indels 86; Gaps 11;

```

QY 169 HPVC-----ODDNENYGRACRDMGKNNFYSGIVDSDGSTFEMKLTSGNVDIY 222
DB 233 HPGLFTFTFESQEMPKESQRMCLKTSEGLPSTR--IKKSKALSFSLQSCRHSIPVF 290
QY 223 --KKLYHS-----DACSSKAVSLRCI-----ACGVNL 248
DB 291 CHSSFYHDTDFLGERLDIYAAKSHRACQKLCITNAVRCQFTYTPAQASCNREGKCYLKL 350
QY 249 NSSR-----OSRIVGESALPGAMPQVSLHYQNV 278
DB 351 SSNGSPFKILHGRGISGYTLRLCKMNECTTKIKPRIVGTAASYRGEMPMQVTLHTTSP 410
QY 279 ---HYCGGSITTPREIVYAAHCVEKPLNPNMHTAFACILROSEF-----FYGAGYQVEK 330
DB 411 TORHLCCGSIIGNMWILTAHCF-YGVESPKILRYSGILNQSEIKEDTSFFG---VOE 465
QY 331 VISHPNYDSTKRNNDIALMLKLOKPLTFNDLVKPYCLPMPGMMLOPEOLCMIWGATTEK 390
DB 466 ITHIDQYKMAESGDIALKLETIVNTDSORPICLPKSGDRNVTYTDCAWTGMYRKLK 525
QY 391 GKTSEVLAAKAVLLIETQRCSRYVDNLITPAMICAGFLOGNVDSCGDSGGPLVTSKN 450
DB 526 DKIONTLQAKAKIPLVTNECCQKRY-RGKHITHKMICAGYREGKDACGDSGGPLSCKN 584
QY 451 NIMWLIGDTSWSGCARKAYRPGYGNVVFPTDWIYROMRA 490
DB 585 EVMHLVGITSWGEGCAQEREPGYVTNVVEYDWILEKTOA 624

```

Search completed: September 26, 2001, 09:27:26
Job time: 316 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 26, 2001, 09:21:45 : Search time 40.65 Seconds
(without alignments)
1601.330 Million cell updates/sec

Title: US-09-615-285-2
Perfect score: 2717
Sequence: 1 MALNCGSPALIGPYENHGY.....YGNVAVPTDVIYRQRADG 492

Scoring table: BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 425026 seqs, 132305027 residues
Total number of hits satisfying chosen parameters: 425026

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: SPREMBL_16:*
2: sp_archaea:*
3: sp_bacteria:*
4: sp_fungi:*
5: sp_human:*
6: sp_invertebrate:*
7: sp_mammal:*
8: sp_mhc:*
9: sp_organelle:*
10: sp_phage:*
11: sp_plant:*
12: sp_rodent:*
13: sp_unclassified:*
14: sp_vertebrate:*
15: sp_virus:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1009	37.1	767	13 Q9DGR2	Q9dgr2 xenopus lae
2	684	25.2	423	4 Q9NZAS	Q9nzas homo sapien
3	676.5	24.9	437	4 Q9NR54	Q9nr54 homo sapien
4	664.5	24.5	445	11 Q9ER03	Q9er03 mus musculu
5	651.5	24.0	457	4 Q9H3S3	Q9h3s3 homo sapien
6	601	22.1	1042	4 Q9Y505	Q9y5q5 homo sapien
7	588	21.6	311	11 Q9ER04	Q9er04 mus musculu
8	587.5	21.6	273	11 Q9ER02	Q9er02 mus musculu
9	577.5	21.3	1113	11 Q9Z319	Q9z319 mus musculu
10	564.5	20.8	855	11 Q9J117	Q9j117 rattus norv
11	563.5	20.7	643	6 Q9Y506	Q9y506 sus scrofa
12	558.5	20.6	855	4 Q9HCA3	Q9hca3 homo sapien
13	554.5	20.4	855	4 Q9H3S0	Q9h3s0 homo sapien
14	544	20.0	1379	5 Q9V4N6	Q9v4n6 drosophila
15	543	20.0	845	13 Q9DGR1	Q9dgr1 xenopus lae
16	533	19.6	571	4 Q9Y495	Q9y495 homo sapien
17	529	19.4	310	11 Q9QY29	Q9qy29 mus musculu
18	527.5	19.4	418	4 Q60235	Q60235 homo sapien
19	522.5	19.2	812	11 Q9R0W3	Q9r0w3 rattus norv

20	509.5	18.8	329	6 Q9GL10	Q9gl10 ovis aries
21	509.5	18.8	389	13 Q9PVX7	Q9pvx7 xenopus lae
22	503.5	18.5	279	11 Q9Q274	Q9q274 rattus norv
23	501	18.4	1524	13 Q91674	Q91674 xenopus lae
24	500.5	18.4	305	11 Q9JH07	Q9jth7 mus musculu
25	499.5	18.4	868	5 Q9Y1V3	Q9y1v3 polyandroca
26	496	18.3	422	4 Q9UL52	Q9ul52 homo sapien
27	493	18.1	307	4 Q9UHW2	Q9uhw2 homo sapien
28	493	18.1	1186	5 Q9YSU2	Q9ysu2 drosophila
29	493	18.1	1449	5 Q9U112	Q9u112 drosophila
30	493	18.1	1462	5 Q9U113	Q9u113 drosophila
31	490.5	18.1	415	6 Q9Z015	Q9z015 sus scrofa
32	489.5	18.0	581	5 Q9XZM7	Q9xzm7 strongyloce
33	473.5	17.4	421	11 Q60491	Q60491 cavia porce
34	469.5	17.3	311	11 Q9QU17	Q9qu17 mus musculu
35	469.5	17.3	312	4 Q9NS34	Q9ns34 homo sapien
36	468.5	17.2	321	4 Q9NR08	Q9nr08 homo sapien
37	468	17.2	297	11 Q88781	Q88781 rattus norv
38	467.5	17.2	321	4 Q9UBR2	Q9ubr2 homo sapien
39	466.5	17.2	806	6 Q18783	Q18783 macropus eu
40	463	17.0	273	6 Q9XSM2	Q9xsm2 ovis aries
41	462	17.0	317	4 Q9GZM4	Q9gzm4 homo sapien
42	460.5	16.9	306	11 Q9ER10	Q9er10 mus musculu
43	459.5	16.9	314	4 Q9Y6M0	Q9y6m0 homo sapien
44	459	16.9	342	11 Q9ES87	Q9es87 rattus norv
45	459	16.9	342	11 Q9ER01	Q9er01 rattus norv

ALIGNMENTS

RESULT 1
Q9DGR2 PRELIMINARY: PRT: 767 AA.
AC Q9DGR2;
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
DE EMBRYONIC SERINE PROTEASE-2.
GN XESP-2.
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidea; Pipidae;
OC Xenopodinae; Xenopus.
OX NCBI_TaxID=8355;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20363741; PubMed=10903452;
RA Yamada K., Takabatake T., Takeshima K.;
RT "Isolation and characterization of three novel serine protease genes
from Xenopus laevis.";
RL Gene 252:209-216(2000).
DR EMBL; AB038497; BAB08217.1; -.
KW Protease.
SQ SEQUENCE 767 AA; 86001 MW; E0566A38796DE96E CRC64;

Query Match 37.1%; Score 1009; DB 13; Length 767;
Best Local Similarity 50.9%; Pred. No. 9.9e-84;
Matches 189; Conservative 49; Mismatches 131; Indels 2; Gaps 2;

QY 120 CGSSGTCINPMSWCDGVSHCGEDENRCVRLYGNETILOVYSQKSKMHPVCCODDWNEN 179
DB 395 CGSSVSCVLSQWCDGVSDCPYGEDMSQSLPADFOLOVSTVSAMLPVCSPTWDD 454
QY 180 YGRACRDMGKYNFNFYS-QGIVDDSGSTSEFMKLTNSAGNVDIYKLYHSDACSSKAVS 238
DB 455 FGRFACQDGRGVSGSSYRNRTDLMSPYAPNGYFKLYSGVWRSKFTSYGVYSSVCYGNVVS 514
QY 239 LRCIACGVNLSSROSRIYGESALPGAMPQVSLAHVONHVCGSIIITEPIVTAACHV 298
DB 515 LHCISGVS-NNSLVSRIYGTGFANLGNMPQVNLQYITGVLCGGSIIISPKWIVTAACHV 573

QY 299 EKLNNPWHMTAFAGILRQSEFMFYAGYQVEKVISHPNYDSKTKNNIDIALKLOKPLTFN 358
 Db 574 YGVSASGMMVHAGTLTKRPSYANASAVEERIIIVHPGYSKYTDNDIALKLRDEITFG 633
 QY 359 DLVPRVCLPNNGMHMQPOLCWSMGATEKTKSEVLNAKAVLLIETORCNSRYVDN 418
 Db 634 YTTQPVCLPNSGMFWEMACTTWTWISGMSGSTYEGSVSTYLOAAIPLIDSNVNCOSYVNG 693
 QY 419 LITPAMICAGFLQGNVDSGQSGPLVTSKNNIMWLIGDTSMGSGCAKAYRPGYGNVM 478
 Db 694 QITSSMICAGTSLGSGVDPCQDSSGGLVKNKNGTWMLVGDTSKMGDCARAKRPGYGNVT 753
 QY 479 VFTDWIYRQMR 489
 Db 754 TFEWISQMR 764

RESULT 2
 Q9NZAS PRELIMINARY: PRT: 423 AA.
 ID Q9NZAS
 AC Q9NZAS;
 DT 01-OCT-2000 (Tremblrel, 15, Created)
 DT 01-OCT-2000 (Tremblrel, 15, Last sequence update)
 DT 01-MAR-2001 (Tremblrel, 16, Last annotation update)
 DE TYPE II MEMBRANE SERINE PROTEASE.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC Smeekens S.S., Lorimer D.D., Wang E., Hou J., Linnevers C.;
 RA "TM-Sp2, a novel type II membrane serine protease expressed in
 RT trachea, colon, and small intestine: identification, cloning, and
 RT chromosomal localization";
 RT Submitted (DEC-1999) to the EMBL/Genbank/DBJ databases.
 RL -1 SIMILARITY: TO CHYMOTRYPSIN SERINE PROTEASE FAMILY (S1).
 CC EMBL: AF216312; AAF31436.1; -;
 DR Interpro: IPR001254; -;
 DR Interpro: IPR001314; -;
 DR Interpro: IPR002172; -;
 DR Pfam: PF00089; Idl_recept_a; 1.
 DR Pfam: PF00089; trypsin; 1.
 DR PRINTS: PRO0722; CHYMOTRYPSIN.
 DR PROSITE: PS00134; TRYPsin_HIS; UNKNOWN_1.
 DR PROSITE: PS00135; TRYPsin_SER; 1.
 DR SMART: SM00192; LDLa; 1.
 KW Protease.
 SQ SEQUENCE 423 AA: 46397 MW: 90792AF0E8AFA30 CRC64;

Query Match 25.2%: Score 684; DB 4; Length 423;
 Best Local Similarity 36.2%: Pred. No. 2.5e-54;
 Matches 165; Conservative 64; Mismatches 167; Indels 60; Gaps 16;
 QY 61 SNPVCTQPKSPSGTCTSKTKALCTLLTGLTFLVGAALAGLMLKRMGSCNSGTEC 120
 Db 2 SNP--CANPSP-WRPSVGIPIITIALSLASTIIIVVLIRKILDKY-----FLC 50
 QY 121 DSSSTCTCINPMWCGVSHCGEGEDENCVRLY--GP-----NFLQYSSQKSMH 169
 Db 51 GQPLHFTPRKQLCGEELDCPLGEDEHCVKSPFPGPAVAVRLSKDRSTLQVLDATGNMF 110
 QY 170 PVCDDNNENYGRAACDMGY--KNFYSQ-----GIYDSSGSTSFMLNTSAGNVD 220
 Db 111 SACDNTTEALAEATRCQMGSSKPTFRRAVEIGPDODLDVETENSELNRRSSG--- 167
 QY 221 IYKRLYSDACSSAAVYSLRACIAGVNLSSRQRIYGGESALPGAMPQVYSLHVQNVH 280
 Db 168 -----PCLSGSLVSLHCLACGKSL---KTPRAVVGGEASVDSMPQVSIQYDKOHV 215
 QY 281 CGSITTPETVTAHCEVKEKLNPNMHTAFAGILR-QSFMFYAGYQVEKVI---SHPN 336
 Db 281 IYKRLYSDACSSAAVYSLRACIAGVNLSSRQRIYGGESALPGAMPQVYSLHVQNVH 280

Db 216 CGGSILDPHWTAAHCEFRKH-TDFENMKVRAGSDKLGSP-----PSLAVAKIIIEFNPM 270
 QY 337 YDSKTKNNIDIALMLOKPLTFNDLVKPCVLPNPGMLQPEOLCWSMGATEEK-GKTS 395
 Db 271 Y---PKDNDIALMLOKPLTFNSGVTRPCLPFPEDELTLPATPLIIGFTKQNGKMSD 327
 QY 396 VLNAKAVLLIETORCNSRYVDNLITPAMICAGFLQGNVDSGQSGPLVTSKNNIMWL 455
 Db 328 ILDAVQVIDISTRCNMADAVQGEYTERKMCAGIPEGVDPCQDSSGGLVKNKNGTWML 386
 QY 456 IGDTSWMSGCAKAYRPGYGNVMFTDWIYRQMRAD 491
 Db 387 VGIVSWGCGGPGSTPGYTRKVSALMVIYMKAE 422

RESULT 3
 Q9NRS4 PRELIMINARY: PRT: 437 AA.
 ID Q9NRS4
 AC Q9NRS4;
 DT 01-OCT-2000 (Tremblrel, 15, Created)
 DT 01-OCT-2000 (Tremblrel, 15, Last sequence update)
 DT 01-MAR-2001 (Tremblrel, 16, Last annotation update)
 DE TRANSMEMBRANE SERINE PROTEASE 3.
 GN TMPS53.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=PANCREATIC CARCINOMA;
 RX MEDLINE=20283276; PubMed=10825129;
 RA Wallerpp C., Hahnel S., Muller-Pillasch F., Burghardt B., Iwanura T.,
 RA Rutenburger M., Lerch M.M., Adler G., Gress T.M.;
 RT "A novel transmembrane serine protease (TMPS53) overexpressed in
 RT pancreatic cancer";
 RL Cancer Res. 60:2602-2606(2000).
 CC -1 SIMILARITY: TO CHYMOTRYPSIN SERINE PROTEASE FAMILY (S1).
 DR EMBL: AF179224; AAF74526.1; -;
 DR Interpro: IPR001254; -;
 DR Interpro: IPR001314; -;
 DR Interpro: IPR002172; -;
 DR Pfam: PF00089; Idl_recept_a; 1.
 DR Pfam: PF00089; trypsin; 1.
 DR PRINTS: PRO0722; CHYMOTRYPSIN.
 DR PROSITE: PS00134; TRYPsin_HIS; UNKNOWN_1.
 DR PROSITE: PS00135; TRYPsin_SER; 1.
 DR SMART: SM00192; LDLa; 1.
 KW Protease.
 SQ SEQUENCE 437 AA: 48204 MW: 351B2FDA8657B12 CRC64;

Query Match 24.9%: Score 676.5; DB 4; Length 437;
 Best Local Similarity 39.1%: Pred. No. 1.3e-53;
 Matches 150; Conservative 57; Mismatches 128; Indels 49; Gaps 13;
 QY 133 CDGVSHCPGEGEDENCVRLY--GP-----NFLQYSSQKSMHNYCQDDNNENY 181
 Db 77 CDGELDPCPLGEDEHCVKSPFPGPAVAVRLSKDRSTLQVLDATGNMFSAFCDFTEALA 136
 QY 182 RAACRDMGY--KNFYSQ-----GIYDSSGSTSFMLNTSAGNVDIYKRLYSDACS 232
 Db 137 ETACRDMGY--KNFYSQ-----GIYDSSGSTSFMLNTSAGNVDIYKRLYSDACS 232
 QY 233 SKAVYSLRACIAGVNLSSRQRIYGGESALPGAMPQVYSLHVQNVHVGSGSITTEPMIV 292
 Db 185 SGLVSLHCLACGKSL---KTPRAVVGGEASVDSMPQVSIQYDKOHVCGSILDPHWT 241
 QY 293 TAAHCEVKEKLNPNMHTAFAGILR-QSFMFYAGYQVEKVI---SHPNYDSKTKNNIDIAL 348
 Db 242 TAAHCEFRKH-TDFENMKVRAGSDKLGSP-----PSLAVAKIIIEFNPM---PKDNDIAL 293
 QY 349 MKLOKPLTFNDLVKPCVLPNPGMLQPEOLCWSMGATEEK-GKTSVNLNAKAVLLIET 407


```

Db      294  MRLQPLFFSGTVRPICLPFDEELTPATPLWIGMGFKONGKMSDILLQASVOIDS 353
OY      408  QMCNRRYVDNLITPRAMICAGFLQGNVDSQDSSGGLVTSKNNIMWLIGDTSWSSGCK 467
Db      354  TRCNADDAVQGEVTKMKCAGIPEGGVDPDCCQDSSGGLVTSKNNIMWLIGDTSWSSGCK 412
OY      468  AYRPGVGNVWFDTMIYROMRAD 491
Db      413  PSTPGVTKVSAVLMYINWKAKE 436

```

RESULT 4

```

OYER03  PRELIMINARY: PRT: 445 AA.
ID 09ER03
AC 09ER03

```

```

DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
DE TYPE 3 SPINESIN.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-BRAIN:
RA Mitsui S., Yamaguchi N.;
RT "CDNA cloning of mouse spinesin."
RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB016230; BAB20277.1; -
SQ SEQUENCE 445 AA; 48502 MW; F60E7CEA6567C7E7 CRC64;

```

```

Query Match      24.5%; Score 664.5; DB 11; Length 445;
Best Local Similarity 35.0%; Pred. No. 1.6e-52;
Matches 145; Conservative 66; Mismatches 176; Indels 27; Gaps 9;

```

```

Db      86  CTTLTGFLVGAALAGLIMKFKMSKCSNSEDSSGTCINPSNWCDDVSHCGGEDE 145
OY      39  CTTLTGFLVGAALAGLIMKFKMSKCSNSEDSSGTCINPSNWCDDVSHCGGEDE 89
OY      146  NRCV-----RLYGNFLLQVYSSQKSWHPYCODMNNENYGRAACRDGY-KNNFY 195
Db      90  BELRLSLPKTVSFRINGEDLLQVLRARPDYLAVCHESWSPALMHIKCKSLGHLRLQH 149
OY      136  SSQGIYVD--SGSISFMKLNYSAGVNDIYKLYHSDACSKAVLSLRTACGVNINSRQ 253
Db      150  KAVNLSIDIKLNSQEFADLSARPGEL-VEEAWKPSANCPSPGRIVSLKSECGAR--PLA 205
OY      254  SRIVGESALPGAMPQVSLHQNHYVCGSITPEMITYTAHCVCK-PLNPNMHTAFA 312
Db      206  SRIVGGQAVASRPMQWQASVMLGSRHGTGASVLAHPWVTAHNCMSFRLSLSSMRVHA 265
OY      313  GLRQSFMEYAGAOVERVISHPNYDSKTKNNDIALMKLQKRLFTNDLYKPYCLPFGMA 372
Db      266  GLVSHGAVRQHQGTWVEKTIIPRLVSAQNHQDVALLDQRTINSDIYGAICLPKAEQY 325
OY      373  LQPEOLCMTSGKATE-EKGTSEVLNAKVLILLETORNSRYVDNLITPRAMICAGFLQ 431
Db      326  PPMWSQCWVSGVGHDPSTHSSDPLQDTMVPILSTHLCNLSGALTHHMLACGYLD 385
OY      432  GNVDSCQDSSGGLVTSKNNIMWLIGDTSWSSGCAKAVRPGVGNVWFDTMIY 485
Db      386  GRADQCGDSSGGLVTSKNNIMWLIGDTSWSSGCAKAVRPGVGNVWFDTMIY 439

```

```

RESULT 5
OYH3S3  PRELIMINARY: PRT: 457 AA.
AC 09H3S3
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)

```

```

DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
DE SPINESIN.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-BRAIN PROSTATE;
RA Mitsui S., Yamaguchi N.;
RT "Molecular cloning of human spinesin."
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB028140; BAB20375.1; -
SQ SEQUENCE 457 AA; 49574 MW; 64406AB4985A2651 CRC64;

```

```

Query Match      24.0%; Score 651.5; DB 4; Length 457;
Best Local Similarity 32.1%; Pred. No. 2.6e-51;
Matches 161; Conservative 63; Mismatches 194; Indels 83; Gaps 16;

```

```

OY      8  PPAIGPYENHGYOPENRYPAQPTVPPYEVNRA-QYRSPVPQYARVYLQASNPVVC 66
Db      9  PPMEOYAE-----EGPQG-----LFRAPGDQDP-----ISQA----- 39
OY      67  TQPKSPSGTVCTSKTKKALCITLTGTFVGAALAGLIMKFKMSKCSN--SGIEDSSG 124
Db      40  -----VCMRSRMRCGAVLGALG-LLAGAGVGSWLLVLYLCRAASPIGTLQDEI 89
OY      125  TCINPSNWCDDVSHCPGGEENRCV-----RLYGNFLLQVYSSQKSWHPYCODD 175
Db      90  TL-----SCSEASAEFALLPALPKTVSFRINSDFLEAQVRDQPRMLLVCHEG 138
OY      176  WNEVNGRAACRDGY-KNNFYSSQGIYVDSSGSTSPKLNKTSAGVNDIYKYL--YNSDA-- 230
Db      139  WSPALGLQICWSLGHRLTHKGVNLTQ-----IKLNSQEFADLSARPGELFEEAMQ 191
OY      231  -----CSSKAVLSLRTACGVNLSNOSRIVGESALPGAMPQVSLHQNHYVCGSIT 286
Db      192  PNNKTSQGVSLKSECGAR---PLASRIYGGQSAVAPRWQASVALGFHTGCGSVL 248
OY      287  TREMITTAHCVCK-PLNPNMHTAFAGILROSFMFYAGVQVERVISHPNYDSKTKND 345
Db      249  APRWVVTAAHCHSPRLARLSSMRVHAGLVSHSAVRPHOGALVERIIPRLVSAQNHQYD 308
OY      346  TALMKLQKRLFTNDLYKPYCLPNNPMLOPEOLCMTSGGATGEEKG-TSEVLNAKVL 404
Db      309  VALLRLQALNFSDTVGAACLPKAEQHPFGKSGRCWVSGVGHHPSTHTYSSDMLQDTVP 368
OY      405  IETORNSRYVDNLITPRAMICAGFLQGNVDSQDSSGGLVTSKNNIMWLIGDTSWSSG 464
Db      369  FSTQLCNSSCVYSGALTPMLCAGYLDGRADACQDSSGGLVCPDQDTRVLYGVVSMGRA 428
OY      465  CAKAVRPGVGNVWFDTMIY 485
Db      429  CAEPNHPGVYAKVAEFLDWIH 449

```

```

RESULT 6
OYI505  PRELIMINARY: PRT: 1042 AA.
AC 09Y505
DT 01-NOV-1999 (TREMBLrel. 12, Created)
DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
DE CORIN.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=HEART;
MEDLINE=99262646; PubMed=10329693;

```

RA Yan W., Sheng N., Seto M., Morser J., Wu Q.;
 RT "Corin, a mosaic transmembrane serine protease encoded by a novel CDNA
 from human heart."
 RL J. Biol. Chem. 274:14926-14935(1999).
 CC -1- SIMILARITY: TO LOW DENSITY LIPOPROTEIN (LDL) RECEPTOR CLASS A
 (LDLRA) DOMAIN.
 CC -1- SIMILARITY: TO SERINE PROTEASES, TRYPSIN FAMILY.
 CC -1- SIMILARITY: TO CHYMOTRYPSIN SERINE PROTEASE FAMILY (S1).
 DR EMBL: AF133845; AAD31850.1; -.
 DR HSSP: P01130; 1A3J.
 DR InterPro: IPR000024; -.
 DR InterPro: IPR001254; -.
 DR InterPro: IPR001314; -.
 DR InterPro: IPR002172; -.
 DR Pfam: PF00057; 1d1_recept_a; 7.
 DR Pfam: PF00089; trypsin; 1.
 DR Pfam: PF01392; Fz_2.
 DR PRINTS: PR00722; CHYMOTRYPSIN.
 DR PRINTS: PR00261; LDLRECEPTOR.
 DR PROSITE: PS01209; LDLRA_1; 6.
 DR PROSITE: PS00068; LDLRA_2; 7.
 DR PROSITE: PS00135; TRYPSIN_SER; 1.
 DR SMART: SM00020; tryp_spc; 1.
 DR GlycoProfile: Hydrobase; Serine protease.
 SQ SEQUENCE 1042 AA; 116564 MW; 7705398BB607AD2 CRC64;

Query Match 22.1%; Score 601; DB 4; Length 1042;
 Best Local Similarity 35.3%; Pred. No. 3e-46;
 Matches 138; Conservative 62; Mismatches 169; Indels 22; Gaps 14;

OY 111 SKCSNGIECDSSGCTGCTINPMWCDVSHCPGGEDENRVRL--YGNPFILOYSSQKRS 167
 DB 653 SFCODELEEC-AHNAVCYRDLMDCEADCDSDSDSEMCDVTLSINWSSSFLMVRHRAE- 710
 OY 168 WHPYCODMNTYGAACRDYKKNFYSSOGIYVDDSGTSFMTKNT---SAGANDIYKK 224
 DB 711 -HHVACADWQIILSQLCKMGLGEPVTKL-IDOEKEPRMLTLHSMESLNGTTLHEL 768
 OY 225 LYHSDGSSKAVVSLRCLAC--CGVNLNSSRSRIYGESALPGAMPQVSLHVO-NVHVC 281
 DB 769 LVNGOSCSRSRKISILCTCKQDGRPARAMNKRILGRTSRGRPMQCSLSQSESGHIC 828
 OY 282 GGSITTPEMVITAAHCVEKPLNPNMHTAFAGT--LRQSFMYGAGYVEKVIHPNYS 339
 DB 829 GCVLIAKKWVTLVAHCFE-GRENAVMKVLGINLIDHPVFMTQRF-VKTIILHPRYSR 886
 OY 340 KTKNNDIALMKLQPLTFNDLVKPYCLPNPGMLOPEQLCMTSGCATEEKGTSEVLNA 399
 DB 887 AVVDISTIVELSEDISETGYRVPCLPNPQWLEPDYCYITGNG--HMGKMPFKLQE 944
 OY 400 AKVLLIETQRCNSRYVDNLITPAMICAGFLQGNVDSQGDGSGPLVTSK-NNIWLIGD 458
 DB 945 GEVAILISLEHQS-YEDMKITTTIRWICAGYSGYVDSQMGDSGGLVEKKGRTLFLGL 1003
 OY 459 TSMGSGC-AKAYRPGVGNVWFTDMYRQM 488
 DB 1004 TSMGSCVCFKVLGPGVSNVSFVEMIKRQI 1034

RESULT 7
 OYER04 PRELIMINARY; PRT; 311 AA.
 AC OYER04;
 DT 01-MAR-2001 (TREMblrel. 16, Created)
 DT 01-MAR-2001 (TREMblrel. 16, Last sequence update)
 DE 01-MAR-2001 (TREMblrel. 16, Last annotation update)
 DE TYPE 2 SPINESIN.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]

RP SEQUENCE FROM N.A.
 RC TISSUE-BRAIN;
 RC Mitsui S., Yamaguchi N.;
 RT "cDNA cloning of mouse spinesin."
 RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AB016229; BAB20276.1; -.
 SQ SEQUENCE 311 AA; 33925 MW; 3124E813A8258F1E CRC64;

Query Match 21.6%; Score 588; DB 11; Length 311;
 Best Local Similarity 40.4%; Pred. No. 1e-45;
 Matches 114; Conservative 44; Mismatches 118; Indels 6; Gaps 4;

OY 206 STSFKMLTSGANDYIKRLYHSDGSSKAVVSLRCLACGVLNLSRSRIRVGESEALPG 265
 DB 28 SQEFAQLSARPGCL-VEAMKRPANCPGRIYSLKSCSGAR---PLASRIYGGQAVASG 83
 OY 266 AMPQVSLHVNHYCGSITTPEMVITAAHCVEK-PLNPNMHTAFAGILRQSFMYGA 324
 DB 84 RWPQASVYMLCSRHTCGASVLAHPWVYTAACHMYSFRLSRLSSRVHAGLVSHGAVRHO 143
 OY 325 GYQVEKVIHPNYSKTKNNDIALMKLQPLTFNDLVKPYCLPNPGMLOPEQLCMTSGW 384
 DB 144 GTMVEKITPPLYSKQNHNDYVALDLQRTPINFSQYDAVCLPAKEQTFPGSCQWISGW 203
 OY 385 GATE-EKGTSEVLNAKAVLLIETQRCNSRYVDNLITPAMICAGFLQGNVDSQGDGSG 443
 DB 204 GHTDPSHTHSSDTLDQDVTWVPLSTHLNCSQWYSGALTRMLCAGLYDGRADACQGDGSG 263
 OY 444 PLYTSKNNIMWLIGDTSWGSCAKAYRPGVGNVWFTDMY 485
 DB 264 PLYCPSGDTWHLVGVSWGRCABRPYAKVAEFLDWIH 305

RESULT 8
 OYER02 PRELIMINARY; PRT; 273 AA.
 AC OYER02;
 DT 01-MAR-2001 (TREMblrel. 16, Created)
 DT 01-MAR-2001 (TREMblrel. 16, Last sequence update)
 DE 01-MAR-2001 (TREMblrel. 16, Last annotation update)
 DE TYPE 1 SPINESIN.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-BRAIN;
 RC Mitsui S., Yamaguchi N.;
 RT "cDNA cloning of mouse type 1 spinesin."
 RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AB016423; BAB20278.1; -.
 SQ SEQUENCE 273 AA; 29662 MW; 63E45FAD2BDF7F5D CRC64;

Query Match 21.6%; Score 587.5; DB 11; Length 273;
 Best Local Similarity 42.2%; Pred. No. 9.8e-46;
 Matches 111; Conservative 40; Mismatches 107; Indels 5; Gaps 3;

OY 225 LYHSDGSSKAVVSLRCLACGVLNLSRSRIRVGESEALPGAMPQVSLHQNHYVCGGS 284
 DB 8 LWSANCPGSRIVSLKSCSGAR---PLASRIYGGQAVASGRWPQASVYMLGSRHTCGAS 64
 OY 285 IITPEMIVTAAHCVEK-PLNPNMHTAFAGILRQSFMYGAGYVEKVIHPNYSKTKN 343
 DB 65 VLAHPWVYTAACHMYSFRLSRLSSRVHAGLVSHGAVRHOQGTWVEKITPPLYSKQNH 124
 OY 344 NDIALMKLQPLTFNDLVKPYCLPNPGMLOPEQLCMTSGGATE-EKGTSEVLNAKY 402
 DB 125 YDVALLDQRTPIINFSQYDAVCLPAKEQTFPGSCQWISGWGHTDPSHTHSSDTLDQDVTW 184
 OY 403 LLIETQRCNSRYVDNLITPAMICAGFLQGNVDSQGDGSGPLVTSKNNIMWLIGDTSWG 462

DB 185 PLSTHLKSSCMYSALHRLMCLACAGYLDGRADACGGSGPLVCSGDTFWHLVGVYVNSG 244
 QY 463 SCGAKAYRPGYGVNMFVTDWY 485
 DB 245 RGAEPNRPYAKVAEFLDWH 267

RESULT 9

Q92319 PRELIMINARY: PRT: 1113 AA.
 AC Q92319;
 DT 01-MAY-1999 (TREMBLrel. 10, Created)
 DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)
 DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
 DE LOW DENSITY LIPOPROTEIN RECEPTOR RELATED PROTEIN 4.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RM SEQUENCE FROM N.A.
 RX MEDLINE-98429596; PubMed-9756624;
 RA Tomita Y., Kim D.-H., Magoori K., Fujino T., Yamamoto T.T.;
 RT "A novel low-density lipoprotein receptor-related protein with type II
 RT membrane protein-like structure is abundant in heart."
 RL J. Biochem. 124:784-789(1998).
 CC -1- SIMILARITY: TO LOW DENSITY LIPOPROTEIN (LDL) RECEPTOR CLASS A
 CC (LDLR) DOMAIN.
 CC -1- SIMILARITY: TO SERINE PROTEASES, TRYPSIN FAMILY.
 CC -1- SIMILARITY: TO CHYMOTRYPSIN SERINE PROTEASE FAMILY (SI).
 DR EMBL: AB013874; BAA34371.1; -.
 DR HSSP: P01130; IAJI.
 DR InterPro: IPR00024; -.
 DR InterPro: IPR001254; -.
 DR InterPro: IPR001314; -.
 DR InterPro: IPR002172; -.
 DR Pfam: PF00057; 1d1_recept_a; 7.
 DR Pfam: PF00089; trypsin; 1.
 DR Pfam: PF01392; Fz; 2.
 DR PRINTS: PRO0722; CHYMOTRYPSIN.
 DR PRINTS: PRO0261; LDLRECEPTOR.
 DR PROSITE: PS01209; LDLRA_1; 6.
 DR PROSITE: PS50068; LDLRA_2; 7.
 DR PROSITE: PS00135; TRYPSIN_SER; 1.
 DR SMART: SM00020; tryp-Spec; 1.
 DR GlycoProtein; Hydrolase; Serine protease.
 SQ SEQUENCE 1113 AA; 122984 MW; B845B2C5F20DD8EC CRC64;

Query Match 21.3%; Score 577.5; DB 11; Length 1113;
 Best Local Similarity 33.3%; Pred. No. 4,6e-44;
 Matches 134; Conservative 67; Mismatches 155; Indels 47; Gaps 14;

QY 111 SKCSNGICDSSGCTCINSNMCDGVSHCPGGEDENRVRLY---GPNFLLQVYSSQKRS 167
 DB 721 SFQDNELFC-AHNECVPRDLMCDGVSDSDDEMGCTLSKNGSSSLTYLTKRAKE- 778
 QY 168 WHPVCDDNMENYGRAACRDMGKNNFYSSQGIYDSDGSTSEFKLNTSAGND---TYKK 224
 DB 779 -HHVCADGRRETLSQLCKOMGLGEP--SVTKLIPQEGCQMLRTPNENLNGSTLQEL 835
 QY 225 LVHSDKSSKAVYSLRCIA--CGVNLSSRSRIYVGESEALPRANWQVSLHQ-VNHYC 281
 DB 836 LVYRHSCPSRSELSILCSKODCGRRPARMKNRIIGRTSPGRWQCSLOSSEPGHIC 895
 QY 282 GGSIIPEPIVTAACVE-----KPLNPMHTAFAGILRQSEFYGAGYQ 327
 DB 896 GCYLIAKKVILVYAHCFEGREDADVKKVYFGINNLDP-----SGFMQTRF----- 941
 QY 328 VEKIVSHPNVDSKTKNNDIALMKLQRPFLTFNDLVKPVCLPNPGMQLQPEOLCMISSWGAT 387
 DB 942 VKTILHPRYSRAVVDYDISVELSDINETSYVRVCLPSPEYLEPPTYCYITGNG-- 999

QY 388 EEKGTSEVLANAKVLLIFTORCNSRYVYDNLITPAMICAGFLQGVDSGCGSGPLTV 447
 DB 1000 HMGKMPFKLQGEVAILPLEOCQS-YEDMKTTTNRMCAGYESGVDSCKMDSGGPLVC 1058
 QY 448 SK-NNIMWLLIGDTSWGSQC-AKAYRPGYGVNMFVTDWYIRQM 488
 DB 1059 ERPGQWTLFGLTSWGSVCSKVLGPGVYSNVSYFGWIERDI 1101

RESULT 10

Q9J17 PRELIMINARY: PRT: 855 AA.
 AC Q9J17;
 DT 01-OCT-2000 (TREMBLrel. 15, Created)
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
 DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
 DE MEMBRANE BOUND SERINE PROTEASE (MEMBRANE BOUND ARGININE SPECIFIC
 DE SERINE PROTEASE).
 GN MBSF.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=101116;
 RN [1]
 RM SEQUENCE FROM N.A.
 RC STRAIN-WISTAR; TISSUE=JEJUNUM;
 RA Tsuzuki S.;
 RT "A membrane bound serine protease expressed in rat small intestine."
 RL Submitted (JAN-2000) to the EMBL/Genbank/DBJ databases.
 RN [2]
 RM SEQUENCE FROM N.A.
 RC STRAIN-WISTAR; TISSUE=DUODENUM;
 RA Inoue H., Takahashi K., Kishi K.;
 RT "membrane-bound arginine specific serine protease."
 RL Submitted (SEP-2000) to the EMBL/Genbank/DBJ databases.
 CC -1- SIMILARITY: TO LOW DENSITY LIPOPROTEIN (LDL) RECEPTOR CLASS A
 CC (LDLR) DOMAIN.
 CC -1- SIMILARITY: TO CHYMOTRYPSIN SERINE PROTEASE FAMILY (SI).
 DR EMBL: AB037898; BAB0502.1; -.
 DR EMBL: AB049189; BAB13765.1; -.
 DR InterPro: IPR000546; -.
 DR InterPro: IPR000859; -.
 DR InterPro: IPR001254; -.
 DR InterPro: IPR001314; -.
 DR InterPro: IPR002106; -.
 DR InterPro: IPR002172; -.
 DR Pfam: PF00057; 1d1_recept_a; 4.
 DR Pfam: PF00089; trypsin; 1.
 DR Pfam: PF00431; CUB; 2.
 DR PRINTS: PRO0722; CHYMOTRYPSIN.
 DR PRINTS: PRO0261; LDLRECEPTOR.
 DR PRODOM: PD00265; -; 1.
 DR PROSITE: PS00339; AA-trNA_LIGASE_II_2; UNKNOWN_1.
 DR PROSITE: PS01180; CUB; 2.
 DR PROSITE: PS01209; LDLRA_1; 2.
 DR PROSITE: PS50068; LDLRA_2; 4.
 DR PROSITE: PS00134; TRYPSIN_HIS; UNKNOWN_1.
 DR PROSITE: PS00135; TRYPSIN_SER; 1.
 DR GlycoProtein; Protease.
 KW VARIANT 665 K -> N.
 SQ SEQUENCE 855 AA; 94955 MW; 35806B7EFC6CF03D CRC64;

Query Match 20.8%; Score 564.5; DB 11; Length 855;
 Best Local Similarity 33.0%; Pred. No. 5,1e-43;
 Matches 128; Conservative 62; Mismatches 131; Indels 67; Gaps 11;

QY 110 GSKCSNGICDSSGCTCINSNMCDGVSHCPGGEDENRVRLYGPFFLLQVYSSQKSMH 169
 DB 522 GCSCPAQSPKRC-SNGKCLPQSQQCKGRDGDGSDASDNNVAVSCYTYTRCQ---N 576
 QY 170 PYCDDMN-ENYGRAACRDMGKNNFYSSQGIYDSDGSTSFMKLNTSAGNVDIYKRLYHS 228

```
Db 577 GCLLKKGNPECDGKKDCSGSDEKNC-----DCGLRSFKR----- 611
Qy 229 DACSSKAVVSLRCLACIGVNLNS-SROSRIYGESALPGAMPQVSLH-VONVHVGGSIT 287
Db 612 -----QARYVGGTNADGEMPMQVSLHALGCGILCASLIS 647
Qy 288 PEWITAAHCVEK-----PLNNPMHTAFAGILROSEFMYGAGYQ---VEKVISHPNYDSK 340
Db 648 PDMIVSAHCFODETIFKYSDDHTMTAFGLGLDQS-KRSASGVGEHKLKRLITTHSPDNF 706
Qy 341 TKNNDIAMLKLOKPLTFNDLYKPVCLPMPGMLOPEOLCWTISGKATEBKCTSEVLNA 400
Db 707 TFDIDIALELEKPEAYESTVARICLPDNTHYFPAKKAIMWGTGHTEGTGALILKKG 766
Qy 401 KVLLETORCSRYVDNLITPAMICAGFLQGNVDSGCGSGPLVT-SKNIMMLIGDT 459
Db 767 EIRVNOTTCER--LLPQOITPRMAGVGFSLGCVDSGCGDSGGLSVYBKDRITQAGV 824
Qy 460 SWGSCAKAYRPGVYGNVWFTDWIYRQ 487
Db 825 SWGEGCAQRNKGYYTRIPEVRDWIKEQ 852

RESULT 11
ID 097506 PRELIMINARY; PRT; 643 AA.
AC 097506;
DT 01-MAY-1999 (Tremblrel. 10, Created)
DT 01-MAY-1999 (Tremblrel. 10, Last sequence update)
DT 01-MAR-2001 (Tremblrel. 16, Last annotation update)
DE KALLIREIN.
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
NCBI_TaxID:9823;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=PLASMA;
RA Takahashi T., Kimura A., Okimura H., Hamabata T.;
RT "Porcine liver plasma kallikrein.";
RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: TO SERINE PROTEASES, TRYPSIN FAMILY.
CC -1- SIMILARITY: TO CHYMOTRYPsin SERINE PROTEASE FAMILY (S1).
DR EMBL; AB022425; BAA37147.1; -.
DR HSSP; P00766; 1CHG.
DR MEROPS; S01.212; -.
DR InterPro; IPR0001177; -.
DR InterPro; IPR001254; -.
DR InterPro; IPR001314; -.
DR InterPro; IPR003014; -.
DR Pfam; PF00089; Trypsin; 4.
DR Pfam; PF00024; PAN; 4.
DR PRINTS; PR00005; APPELOMAIN.
DR PRINTS; PR00722; CHYMOTRYPsin.
DR PROSITE; PS00495; APPEL; 4.
DR PROSITE; PS00134; TRYPSIN_HIS; UNKNOWN_1.
DR PROSITE; PS00135; TRYPSIN_SER; UNKNOWN_1.
DR SMART; SM00020; TRYP_SPC; 1.
DR KMW; KMW00000; Serine protease.
SQ SEQUENCE 643 AA; 72227 MW; AFF2923E3C3CB80A CRC64;

Query Match 20.7%; Score 563.5; DB 6; Length 643;
Best Local Similarity 29.9%; Pred. No. 4,5e-43;
Matches 154; Conservative 63; Mismatches 185; Indels 113; Gaps 19;
```

```
Db 201 PCADSEICHHMDFIOLHAFSDVDVARIAPDAFV-----C-----RTICTYHPNCLF 248
Qy 160 VY-----SSQR-----KSMH-----PVCODDMNENYGRAACRDM---GYKNFYSSQ 198
Db 249 TETVAMKIEGSRNNCFIKTSHSGTSPFTPEMNAISGTSLLTQOTLPEDCHSKITYSEV 308
Qy 199 GIYVDSGTSFEM-----KLNTSAGNVDIYKKLYSDAC-SSKAVVSLR----- 240
Db 309 DFEQELNVTFOVAGANLCOEFTCTKTRICQFPTYSLHPEDCGKEKCKSLRSSDGPYKI 368
Qy 241 -----C-----IACGVNLNSRQSRIVGESALPGAMPQVSLH-VONVHVGGSIT 287
Db 369 THGRASSGYSLRLDRSGDHSACATKAN---TRIVGTDLSLGEPMQVSLHALGCGILCASLIS 647
Qy 278 VHVCGSITPEWITAAHCVEKPLNNPMHTAFAGILROSEFMYGAGYQ-OVEKVISHPN 336
Db 425 -HLCGSGIIGHQWVLTAAHCFD-GSLPDIRIRYIGLINISEITKEPFSQVKEIILHN 482
Qy 337 YDSKTKNNDIAMLKLOKPLTFNDLYKPVCLPMPGMLOPEOLCWTISGKATEBKCTSEV 396
Db 483 KYIESGHDIALLKLEPLNTDFEQPICLPSRDDTNVYVYNCWVTGFTGEKEGEIONI 542
Qy 397 LNAKVLLETORCSRYVDNLITPAMICAGFLQGNVDSGCGSGPLVT-SKNIMMLIGDT 459
Db 543 LQKNVPIPLVSNBECQKST-RDKKISKQMICAGYKGGKDGSGGLVCKYKINGIMHLV 601
Qy 457 GDTWSGSCAKAYRPGVYGNVWFTDWIYRQMRAD 491
Db 602 GTTWSGEGCARREGPGYVTVKIEYMDWILEKTQDD 636

RESULT 12
ID 09HCA3 PRELIMINARY; PRT; 855 AA.
AC 09HCA3;
DT 01-MAR-2001 (Tremblrel. 16, Created)
DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
DT 01-MAR-2001 (Tremblrel. 16, Last annotation update)
DE SERINE PROTEASE TADG15.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
NCBI_TaxID:9606;
RN [1]
RP SEQUENCE FROM N.A.
RC Tissue=Brain;
RA O'Brien T.J.;
RT Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF057145; AAG15395.1; -.
DR KMW; KMW00000; Serine protease.
SQ SEQUENCE 855 AA; 94700 MW; 08827280FA2D2BF3 CRC64;

Query Match 20.6%; Score 558.5; DB 4; Length 855;
Best Local Similarity 33.9%; Pred. No. 1.8e-42;
Matches 131; Conservative 57; Mismatches 129; Indels 69; Gaps 14;
```


DR PRINTS: PRO0722: CHYMOTRYPSIN.
 DR PRINTS: PRO0261: LDLRECEPTOR.
 DR PROSITE: PS01209: LDLRA_1; 1.
 DR PROSITE: PS50068: LDLRA_2; 2.
 DR PROSITE: PS50287: SRCR_2; 1.
 DR PROSITE: PS00134: TRYPSIN_HIS; 1.
 DR PROSITE: PS00135: TRYPSIN_SER; 1.
 DR PROSITE: PS00227: TUBULIN; 1.
 DR SMART: SM00020: Tryp_Spc; 1.
 KW GTP-binding: Glycoprotein; Hydrolase; Microtubules; Serine protease.
 SQ SEQUENCE 1379 AA; 149489 MW; A593A9CC2167EAB7 CRC64;

Query Match 20.0%; Score 544; DB 5; Length 1379;
 Best Local Similarity 33.4%; Pred. No. 7e-41;
 Matches 141; Conservative 65; Mismatches 158; Indels 58; Gaps 20;

QY 111 SKCSNSG---IECDSSGTCINSMNCDGSHCPGGEDENRCLYGPN-----FILQVYS 162
 Db 922 AKCEGCGPEITCGDS-QCIGTKHICDGIIDCPYGDENRCLSRNEDVGTGVLEVIR 980
 QY 163 SQRKSMHPVQCDDMNENYGRACRDMGYKNFYSSOGIVDDSGSTSPFK-----LNT 214
 Db 981 IGOBQWMPACVKNMORAVSPASVCSILGY-----SAVNATSVLTQLTHRPLAT 1029
 QY 215 SAGNDIYK-----KLYHSDACSSKA-----VSLRC--IACG-VNLNSSRQS-RIVGG-259
 Db 1030 VNVSTIDIKMKVAKRRKSTLMQEPANCKTEDYPMADLTCSNYCGRKRRHRSRRIIG 1089
 QY 260 ESALPGAMPQVSL--HYQNVHVGCGSITTPEMIVTAACHV-EKPLNPNMHTAFAGILR 316
 Db 1090 TQASGNNPFLAIIIGEPKITYCAGVILSDQWVLTASHCVNTSYVIDEDMTIQLGYTR 1149
 QY 317 Q-SFMFYAGYQVERVYISHPNYD-SKTKNNDIALMKLQKPLTFNDLVKPVCLPBNQM-L 373
 Db 1150 RNSFYSGQKVKVKAIVHPQYMAIADNDIALFQLATRAVFHEHLLPVCLPSPSVRL 1209
 QY 374 QPEQLCMISGKATEBEKKTSE---VLNAKAVLLLETQRCNSRYVDNL-TTPAMICAGF 429
 Db 1210 HPGTLCITYIGWCKREDKDKPKSTYEYIVNEVOYPIITRNQCD--WLDNLTYSSEGVACGF 1267
 QY 430 LOGNVDSGCGDSGPLY--TSKNNIWMLIGDTSWGSAGCAKAYRPGVGNVWFTDWIR 486
 Db 1268 DDGGADACGDSGPLYLPYBEKRNRFVGGIVSMGIMCAHRLPGVIANVOYPIWIOE 1327
 QY 487 QM 488
 Db 1328 QI 1329

RESULT 15
 Q9DGR1 PRELIMINARY; PRT; 845 AA.
 AC Q9DGR1;
 DT 01-MAR-2001 (Tremblrel. 16, Created)
 DT 01-MAR-2001 (Tremblrel. 16, last sequence update)
 DT 01-MAR-2001 (Tremblrel. 16, last annotation update)
 DE HOMOLOG OF HUMAN MT-SPL.
 GN XMT-SPL.
 OS Xenopus laevis (African clawed frog).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipridae;
 OC Xenopodidae; Xenopus.
 OC NCBI_TaxID=8355;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=20363741; PubMed=10903452;
 RA Yamada K., Takabatake T., Takeshima K.,
 RT Isolation and characterization of three novel serine protease genes
 RT from Xenopus laevis.
 RL Gene 252:209-216(2000).
 DR EMBL: AB038498; BAB08218.1;
 SQ SEQUENCE 845 AA; 93597 MW; 7FD7E62851A758B7 CRC64;

Query Match 20.0%; Score 543; DB 13; Length 845;
 Best Local Similarity 27.9%; Pred. No. 4.7e-41;
 Matches 136; Conservative 72; Mismatches 173; Indels 106; Gaps 12;

QY 20 YQPEHPYPAQPTVPTVYEVHQAQYPPSPVQYARVLTQASNPVYCPQSPSPSTVCTS 79
 Db 435 YEPHNPQDQ-----FTCRSGRCIRIDOKCDGNDED 467
 QY 80 KTKKALCITLTGTFVGAALA-----AGLLMKFPGSKNSGTEDSSGTCI 127
 Db 468 FSDMSCTCTALQFCVNSKCLKPSTFYICDGVNDCGSSDELACCPNTRFC-GNGKCI 526
 QY 128 NPSNCDGVSHPGGEDENRCLYGPNFILQVYSSQRKSMHPVQCDDMNENYGRACRD 187
 Db 527 PDSQCDRYDNCGDSDEAECDQVL-----TTACTE 557
 QY 188 MKYKNFYSSOGIVDDSGSTSPFKLNTSAGNDIYKLYHSDA---GSKAVVSLRCIAC 244
 Db 558 YTYK-----CKNNOCITRKKNPECDGSDGSDENAKCNC 594
 QY 245 GVNLSRSRQRIYGGESALPGAMPQVSLHYQ-NVHVGCGSITTPEMIVTAACHVEKP-- 301
 Db 595 G-KRPFTKSRIVGVNADTGEFPMQVSLHAKGNHTCGASLGFTMLISAHCFQDDHQ 653
 QY 302 --LNNPWHMTAFAGILRQSFMYG--AGYQVERVISHPNYDSTKKNNDIALMKLQKPLTF 357
 Db 654 MRYSDASLMTAYLGLHDQALNTKDVVERIRIKRIAHIGHFNNTYDNDIAVLELEKPYEY 713
 QY 358 NDLVAPVCLPBNQMMLQPEQLCMISGKATEBEKKTSEVLNAKAVLLLETQRCNSRYVD 417
 Db 714 TDFIQPCIPSTHDPVYKPLIWTGALKEGGAAVILQKAEIRIINOTECNK--LLD 771
 QY 418 NLITPAMICAGFLOGNVDSGCGDSGPLYTSK-NNIWMLIGDTSWGSAGCAKAYRPGVGN 476
 Db 772 GQLTPMILCAGVSGSIDACGDSGGLSSVETLNKYLIVAGVSMGECARNNRPVYTK 831
 QY 477 VNVFTDW 483
 Db 832 VSMRDM 838

Search completed: September 26, 2001, 09:27:03
 Job time: 318 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OW protein - protein search, using sw model

Run on: September 26, 2001, 09:22:49 : Search time 34.76 Seconds
(without alignments)
415.089 Million cell updates/sec

Title: US-09-615-285-2_COPY_255_492
Perfect score: 1319
Sequence: 1 RIVGGSALPGAMPQVSLH.....VYGNVVFETDWYRQWRADG 238

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 412676 seqs, 60623988 residues

Total number of hits satisfying chosen parameters: 412676

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

A_Geneseq_0601:*

1: /SIDSR/gcgdata/geneseq/geneseq/AA1980.DAT:*
2: /SIDSR/gcgdata/geneseq/geneseq/AA1981.DAT:*
3: /SIDSR/gcgdata/geneseq/geneseq/AA1982.DAT:*
4: /SIDSR/gcgdata/geneseq/geneseq/AA1983.DAT:*
5: /SIDSR/gcgdata/geneseq/geneseq/AA1984.DAT:*
6: /SIDSR/gcgdata/geneseq/geneseq/AA1985.DAT:*
7: /SIDSR/gcgdata/geneseq/geneseq/AA1986.DAT:*
8: /SIDSR/gcgdata/geneseq/geneseq/AA1987.DAT:*
9: /SIDSR/gcgdata/geneseq/geneseq/AA1988.DAT:*
10: /SIDSR/gcgdata/geneseq/geneseq/AA1989.DAT:*
11: /SIDSR/gcgdata/geneseq/geneseq/AA1990.DAT:*
12: /SIDSR/gcgdata/geneseq/geneseq/AA1991.DAT:*
13: /SIDSR/gcgdata/geneseq/geneseq/AA1992.DAT:*
14: /SIDSR/gcgdata/geneseq/geneseq/AA1993.DAT:*
15: /SIDSR/gcgdata/geneseq/geneseq/AA1994.DAT:*
16: /SIDSR/gcgdata/geneseq/geneseq/AA1995.DAT:*
17: /SIDSR/gcgdata/geneseq/geneseq/AA1996.DAT:*
18: /SIDSR/gcgdata/geneseq/geneseq/AA1997.DAT:*
19: /SIDSR/gcgdata/geneseq/geneseq/AA1998.DAT:*
20: /SIDSR/gcgdata/geneseq/geneseq/AA1999.DAT:*
21: /SIDSR/gcgdata/geneseq/geneseq/AA2000.DAT:*
22: /SIDSR/gcgdata/geneseq/geneseq/AA2001.DAT:*

Pred. NO. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1319	100.0	492	AAV92050	HRPc6/7 polypept1
2	1319	100.0	492	AAV77726	Human tumour supp
3	1319	100.0	492	AAV44406	Human 20P1P12-GTC2
4	1319	99.3	283	AAV81492	Human prostate-ass
5	1303	98.8	492	AAAB36901	Human TMPRSS2 prot
6	1303	98.8	492	AAV57280	Ovrl15 homolog pro
7	655	49.7	452	AAV41694	Human PRO382 prote
8	655	49.7	454	AAAB32246	Tumour associated
9	654	49.6	453	AAAB44250	Human PRO382 (UNQ3
10	645.5	48.9	248	AAAB3572	Human cancer assoc
11	645.5	48.9	327	AAV72093	Human serine prote

12	564	42.8	414	21	AAAB08912	Human secreted pro
13	564	42.8	480	21	AAAB08950	Human secreted pro
14	558	42.3	296	21	AAV72108	Human serine prote
15	558	42.3	372	21	AAV72092	Human serine prote
16	558	42.3	457	21	AAAB1699	Human serine prote
17	549	41.6	273	21	AAAB1696	Mouse serine prote
18	549	41.6	311	21	AAAB1697	Mouse serine prote
19	549	41.6	445	21	AAAB1698	Mouse serine prote
20	544	41.2	238	21	AAAB1695	Mouse serine prote
21	523.5	39.7	798	15	AAV57283	Bovine enterokinase
22	521	39.5	432	22	AAV99417	Human PRO1570 (UNQ
23	521	39.5	432	22	AAAB6166	Human PRO1570. Ho
24	521	39.5	432	22	AAAB6166	Protein of the inv
25	521	39.5	435	20	AAV06437	Human protease HUP
26	521	39.5	435	22	AAV72558	Human seripancrin
27	518.5	39.3	418	17	AAAB9435	Trypsin-like enzym
28	518.5	39.3	418	20	AAV29498	Human lung tumour
29	518.5	39.3	418	20	AAV29501	Human lung tumour
30	518.5	39.3	418	20	AAV29502	Human lung tumour
31	518.5	39.3	418	21	AAAB4428	Human lung tumour
32	518.5	39.3	418	21	AAAB4437	Human lung tumour
33	518.5	39.3	418	21	AAAB4438	Human lung tumour
34	516	39.1	233	18	AAAB2986	Human seripancrin
35	516	39.1	492	22	AAV72559	Trypsin-like enzym
36	513.5	38.9	232	17	AAAB9430	Mouse hepsin prote
37	507.5	38.5	416	20	AAV6812	A mouse serine pro
38	507.5	38.5	416	20	AAAB8395	Complete mouse pla
39	502.5	38.1	812	16	AAAB3959	Murine plasminogen
40	502.5	38.1	812	17	AAW07585	Murine plasminogen
41	502.5	38.1	812	20	AAV08686	Murine plasminogen
42	502.5	38.1	812	20	AAV94036	Murine plasminogen
43	502.5	38.1	812	21	AAAB0053	Murine plasminogen
44	502.5	38.1	812	21	AAAB16490	Murine plasminogen
45	502	38.1	683	21	AAAB19551	Human matrilysin (

ALIGNMENTS

RESULT 1

AAV92050

ID AAV92050 standard; Protein: 492 AA.

XX

AC AAV92050;

XX

DT 01-AUG-2000 (first entry)

XX

DE HRPc6/7 polypeptide from androgen-inducible gene clone.

XX

KW Androgen inducible; testosterone; prostate cancer; cytostatic;

KW TMRRS2; diagnosis.

XX

OS Homo sapiens.

XX

PN WO200018961-A2.

XX

PD 06-APR-2000.

XX

PE 30-SEP-1999; 99WO-US22535.

XX

PR 30-SEP-1998; 98US-0163759.

XX

PR 30-SEP-1998; 98US-0164159.

XX

PA (MILL-) MILLENNIUM PHARM INC.

XX

PI Macbeth KJ, Shyjan AW;

XX

DR WPI; 2000-293182/25.

XX

DR N-PSDB; AAA08603.

XX

PT Novel methods for identifying compounds for treating prostate cancer

PT comprising measuring the level of expression or activity of 1 or more

PT of 11 genes or their products

```
XX PS Claim 2; Fig 3; 108bp; English.
XX
CC This protein is encoded by a gene which is androgen (e.g. testosterone)
CC inducible in androgen-dependent prostate cancer cells (e.g. LNCap cells)
CC and constitutively expressed in androgen-independent prostate cancer
CC cells (e.g. LNCap cells). Agents which decrease the expression or
CC activity of these clones may slow or arrest the growth of prostate cancer
CC cells or may kill them. HRPca6/7 can be obtained from the sequence of
CC the known gene for TMPRSS2. A compound useful for treating prostate
CC cancer can be identified in a novel method comprising measuring the
CC expression level, or activity, of HRPca2, 3, 6/7, 8, 9, 10, 13, 14, 15,
CC 19, or peripherin-type benzodiazepine receptor (PBR) in a cell, in the
CC presence and absence of a test compound. The sequences may also be used
CC in diagnosis of prostate cancer and to determine efficacy of treatment
CC for prostate cancer.
XX
SQ Sequence 492 AA;

Query Match          100.0%; Score 1319; DB 21; Length 492;
Best Local Similarity 100.0%; Pred. No. 5.4e-117;
Matches 238; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 RIVGESALPGAMPWOVSLHVQNVHVGSGSITTPEMWYTAACHVEKPLNPMWHTAFAGI 60
DB 255 rlvgsesalpgawpwyshvqnvhvcgslltpewltaahcveklpnpwhwtatag1 314
OY 61 LROSFMYGAGYOEKYSHPNYSKTRKNDIALMKLOKPLTFNDLYKPVCLPMPGMQLQ 120
DB 315 lrgsfmlygagyekevshpnysktrkndialmk1qkpltfndlykpvclpmpgmmlq 374
OY 121 PEOLCMTISGMCATEEKGKTSSEVLNAKAVLLIETORCNSRYVYDNLITPAMICAGFLQGNV 180
DB 375 peqlcwtisgwateekgktssevlnaakvllletqrcnsryyddnlitpamlcagflqgnv 434
OY 181 DSCGDSGSGPLVTSKNNIMWLIGDTSWGSACAKAYRPGVYGVNMFDMWIRQMRADG 238
DB 435 dscgdsdsgplvtlsknnimwllgdtswsgsacakayrpgvygvnmvfdmlyrqrnadg 492

RESULT 2
AAV77726
ID AAV77726 standard; Protein: 492 AA.
XX
AC AAV77726;
XX
DT 12-MAY-2000 (first entry)
XX
DE Human tumour suppressor TMPRSS2 polypeptide.
XX
KW Tumour suppressor gene; TMPRSS2; cancer; human; drug design;
KM gene therapy; protein therapy.
XX
OS Homo sapiens.
XX
PN WO200000605-A1.
XX
PD 06-JAN-2000.
XX
PF 29-JUN-1999; 99WO-US14622.
XX
PR 29-JUN-1998; 98US-0091044.
XX
PA (MYRI-) MYRIAD GENETICS INC.
XX
PI Mong AKC, Tavtliglan SV, Teng DHF;
XX
DR WPI: 2000-170914/15.
XX
DR N-PSDB; AA287786.
XX
PT Novel tumor suppressor TMPRSS2 used for the diagnosis and prognosis of
PT human cancer -
```

```
XX PS Claim 55; Page 77-79; 89pp; English.
XX
CC The invention provides a new tumour suppressor gene, designated TMPRSS2.
CC The TMPRSS2 polynucleotides and polypeptides can be used in methods for
CC diagnosing and prognosing predisposition to cancer in humans. The
CC polypeptides may also be used in assays to screen for compounds with
CC anti-cancer or therapeutic properties. The polypeptides are also useful
CC for rational drug design. The TMPRSS2 polynucleotides and polypeptides
CC may be used for gene therapy and protein therapy. The present sequence
CC represents the TMPRSS2 polypeptide.
XX
SQ Sequence 492 AA;

Query Match          100.0%; Score 1319; DB 21; Length 492;
Best Local Similarity 100.0%; Pred. No. 5.4e-117;
Matches 238; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 RIVGESALPGAMPWOVSLHVQNVHVGSGSITTPEMWYTAACHVEKPLNPMWHTAFAGI 60
DB 255 rlvgsesalpgawpwyshvqnvhvcgslltpewltaahcveklpnpwhwtatag1 314
OY 61 LROSFMYGAGYOEKYSHPNYSKTRKNDIALMKLOKPLTFNDLYKPVCLPMPGMQLQ 120
DB 315 lrgsfmlygagyekevshpnysktrkndialmk1qkpltfndlykpvclpmpgmmlq 374
OY 121 PEOLCMTISGMCATEEKGKTSSEVLNAKAVLLIETORCNSRYVYDNLITPAMICAGFLQGNV 180
DB 375 peqlcwtisgwateekgktssevlnaakvllletqrcnsryyddnlitpamlcagflqgnv 434
OY 181 DSCGDSGSGPLVTSKNNIMWLIGDTSWGSACAKAYRPGVYGVNMFDMWIRQMRADG 238
DB 435 dscgdsdsgplvtlsknnimwllgdtswsgsacakayrpgvygvnmvfdmlyrqrnadg 492

RESULT 3
AAV44406
ID AAV44406 standard; Protein: 492 AA.
XX
AC AAV44406;
XX
DT 22-MAR-2000 (first entry)
XX
DE Human 20P1F12-GTC2 protein.
XX
KW 20P1F12; TMPRSS2; androgen; serine protease; 20P1F12-GTC1; cancer;
KM transmembrane protein; colon; prostate; prostate tumour.
XX
OS Homo sapiens.
XX
PN WO9962942-A2.
XX
PD 09-DEC-1999.
XX
PF 01-JUN-1999; 99WO-US12253.
XX
PR 01-JUN-1998; 98US-0087598.
XX
PR 29-JUN-1998; 98US-0091474.
XX
PR 14-APR-1999; 99US-0129521.
XX
PA (UROG-) UROGENESIS INC.
XX
PA (AFAR/) AFAR D E.
XX
PA (HUBE/) HUBERT R S.
XX
PA (LEON/) LEONG K.
XX
PA (RAIT/) RAITANO A B.
XX
PA (SAFF/) SAFFRAN D C.
XX
PI Afar DE, Hubert RS, Leong K, Raitano AB, Saffran DC;
XX
DR WPI: 2000-116363/10.
XX
DR N-PSDB; AA229636.
XX
```

PF Novel cell surface antigen useful to treat colon and prostate cancer -
 XX
 PS Claim 1; Fig 1; 58pp; English.
 XX

CC The present sequence is the 20P1F12 protein (also known as the TMPRSS2
 CC protein) which is a prostate-specific, androgen-regulated, cell surface
 CC serine protease. It is a glycosylated type II transmembrane protein with
 CC an extracellular C-terminal serine protease domain, a scavenger receptor
 CC cysteine-rich domain, an LDL receptor-class A domain and a predicted
 CC transmembrane domain. Host cells can be transformed to produce this
 CC protein, using vector containing 20P1F12/TMPRSS2 gene (also designated
 CC 20P1F12-G1C1, as deposited with ATCC accession number 207097).
 CC Anti-20P1F12/TMPRSS2 antibodies may be used as therapeutic agent for
 CC prostate and colon cancers, to image prostate cancer cells and
 CC prostate tumours, to identify ligands and cellular constituents that
 CC bind to a 20P1F12/TMPRSS2 gene product and for use as cancer vaccines.
 XX

XX Sequence 492 AA:

Query Match 100.0%; Score 1319; DB 21; Length 492;
 Best Local Similarity 100.0%; Pred. No. 5.4e-117; Indels 0; Gaps 0;
 Matches 238; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RIVGESALPGAMPQVSLHVQNVHVCSSITPEWIVTAACVEKPLNNPMTAFAGI 60
 DB 255 rivgesalpgampqvslhvqnvhvcgsiltpewivtaahcveklpnnpmtafagi 314
 QY 61 LRQSFMEYGAGYQVEKVISHPNYSKTKNDIALMKLQKPLTFNDLVKPCLPNGMILQ 120
 DB 315 lrqsfmefyagayqvekvishpnysktnndialmklqkpltfndlvkpcclpnpgmilmq 374
 QY 121 PEOLCWISMGATFEKGTSEVINAQVLLIETQRNSRYVDNLITPMICAGFLQGNV 180
 DB 375 peqlcwiswgateekgtsevinakvllietqrnsryvydniltpmicagflqgnv 434
 QY 181 DSCQGDGSGPLVTSKNNIMWLIGDTSMSGSCAKAYRPGYGVNMVFTDVIYROMRADG 238
 DB 435 dscqgdsgsplvtsknnimwlligdtswsgscakayrpgygvnmvftdviyqmrtdg 492

RESULT 4
 AAY81492
 ID AAY81492 standard; Protein; 283 AA.
 XX
 AC AAY81492;
 XX
 DT 18-JUL-2000 (first entry)
 XX
 DE Human prostate-associated protease (HUPAP).
 XX
 KW Human prostate-associated protease; HUPAP; kallikrein; serine protease;
 KW gastrointestinal disorder; cancer; prostate disorder.
 XX
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT Modified-site 4 /note= "N-glycosylated"
 FT Misc-difference 235 /label= unknown
 FT /note= "Encoded by CNT"
 XX
 PN US6043033-A.
 XX
 PD 28-MAR-2000.
 XX
 PE 27-FEB-1997; 97US-0807151.
 XX
 PR 27-FEB-1997; 97US-0807151.
 XX
 PA (INCY-) INCYTE PHARM INC.
 XX

PI Bandman O, Lal P;
 XX
 DR WPI: 2000-28523/24.
 DR N-PSDB: AAA12975.
 XX

PT Polynucleotide encoding human prostate-associated protease useful for
 PT diagnosing and treating cancers, prostate disorders and
 PT gastrointestinal disorders -
 XX
 PS Claim 1; Fig 1A-B; 27pp; English.
 XX

CC This sequence represents human prostate-associated protease (HUPAP). cDNA
 CC encoding HUPAP was initially identified in a spinal cord cDNA library,
 CC the cDNA encoding this sequence representing a consensus of overlapping
 CC and/or extended nucleic acid sequences from spinal cord, prostate tumour
 CC and colon cDNA libraries. HUPAP is a serine protease with structural and
 CC functional homology with bovine enterokinase, human pancreatic kallikrein
 CC and African rat renal kallikrein, sharing 38% homology with bovine
 CC enterokinase. In addition, HUPAP is related to prostate-specific antigen
 CC (PSA), a kallikrein which is a highly sensitive marker for prostate
 CC cancer. HUPAP and nucleic acids encoding it are useful for the diagnosis,
 CC prevention and treatment of gastrointestinal disorders such as ulcerative
 CC colitis, pancreatitis, cancers, and prostatic disorders. HUPAP agonists
 CC may be used to treat gastrointestinal disorders, and HUPAP antagonists
 CC and inhibitors may be used to suppress excessive cell proliferation,
 CC which is of use in cancer therapy. HUPAP nucleic acids are also useful
 CC for generating hybridisation probes which may be used for mapping
 CC naturally occurring genomic sequences.
 XX

XX Sequence 283 AA:

Query Match 99.3%; Score 1310; DB 21; Length 283;
 Best Local Similarity 99.6%; Pred. No. 2e-116; Indels 1; Gaps 0;
 Matches 237; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 RIVGESALPGAMPQVSLHVQNVHVCSSITPEWIVTAACVEKPLNNPMTAFAGI 60
 DB 46 rivgesalpgampqvslhvqnvhvcgsiltpewivtaahcveklpnnpmtafagi 105
 QY 61 LRQSFMEYGAGYQVEKVISHPNYSKTKNDIALMKLQKPLTFNDLVKPCLPNGMILQ 120
 DB 106 lrqsfmefyagayqvekvishpnysktnndialmklqkpltfndlvkpcclpnpgmilmq 165
 QY 121 PEOLCWISMGATFEKGTSEVINAQVLLIETQRNSRYVDNLITPMICAGFLQGNV 180
 DB 166 peqlcwiswgateekgtsevinakvllietqrnsryvydniltpmicagflqgnv 225
 QY 181 DSCQGDGSGPLVTSKNNIMWLIGDTSMSGSCAKAYRPGYGVNMVFTDVIYROMRADG 238
 DB 226 dscqgdsgsplvtsknnimwlligdtswsgscakayrpgygvnmvftdviyqmrtdg 283

RESULT 5
 AAB36901
 ID AAB36901 standard; Protein; 492 AA.
 XX
 AC AAB36901;
 XX
 DT 26-FEB-2001 (first entry)
 XX
 DE Human TMPRSS2 protein.
 XX
 KW Prostate specific androgen regulated protein; ARSDR1; TMPRSS2;
 KW PART-1; neoplastic.
 XX
 OS Homo sapiens.
 XX
 PN WO200065067-A2.
 XX
 PD 02-NOV-2000.
 XX
 PF 21-APR-2000; 2000WO-US10920.
 XX

XX 23-APR-1999; 99US-0130778.
PR 30-AUG-1999; 99US-0151585.
PR 30-DEC-1999; 99US-0174003.
PR 24-JAN-2000; 2000US-0177751.
XX
PA (UNIW) UNIV WASHINGTON.
XX
PI Nelson PS, Hood L, Lin B;
XX
DR WPI: 2000-679676/66.
DR N-PSDB; AAC83325.
XX
PT Polynucleotide encoding prostate specific androgen regulated
PT peptides and inhibitor of the peptides useful for treating or
PT reducing the progression of prostate neoplastic condition in an
PT individual -
XX
PS Claim 63; Page 86-88; 121pp; English.
XX
PS The present invention relates to prostate specific androgen regulated
CC proteins. The invention may be used to determine an expression level
CC of the prostate-specific proteins ARSDR, TMPRSS2, or PART-1 in a
CC fluid sample or prostate cell sample from an individual. It may also
CC be used for diagnosing and predicting the susceptibility of a
CC prostate neoplastic condition in an individual. Inhibitors of the
CC proteins are useful for treating or preventing the progression of a
CC prostate neoplastic condition.
XX
SQ Sequence 492 AA;

Query Match 98.8%; Score 1303; DB 21; Length 492;
Best Local Similarity 98.3%; Pred. No. 1.8e-115;
Matches 234; Conservative 3; Mismatches 1; Indels 0; Gaps 0;
QY 1 RIVGESALPGAMPMQVSLHVNQVHVCSSITTPBWIYTAHCVEKPLNPNHMTAFAGI 60
DB 255 RIVGESALPGAMPMQVSLHVNQVHVCSSITTPBWIYTAHCVEKPLNPNHMTAFAGI 314
QY 61 LROSFMFAGAGYQVQKVSHPNYSKTKNDIALMKLQKPLTFNDLVKPVCLPDPGMMLQ 120
DB 315 LRSFMTYGAQYQVQKVSHPNYSKTKNDIALMKLQKPLTFNDLVKPVCLPDPGMMLQ 374
QY 121 PEOLCWSIGMGATEEKGKTSVLYNAKVLLETQRCNSRYVDNLITPAMICAGFLQGNV 180
DB 375 PEQLCWSIGMGATEEKGKTSVLYNAKVLLETQRCNSRYVDNLITPAMICAGFLQGNV 434
QY 181 DSCGDSGGLPVTSTKNNIWMILIGDTSMGSCAKATRPVYGVNMFDTMTIRQMRADG 238
DB 435 DSCGDSGGLPVTSTKNNIWMILIGDTSMGSCAKATRPVYGVNMFDTMTIRQMRADG 492

RESULT 6
AA57280
ID AA57280 standard; Protein: 492 AA.
XX
AC AA57280;
XX
DT 06-JUN-2000 (first entry)
XX
DE Ovr115 homolog protein.
XX
KW CS6; cancer specific gene; cancer; gynecologic cancer; ovarian; breast;
KW endometrial; uterine; lung; cytotoxic.
XX
OS Homo sapiens.
XX
PN WO200012758-A1.
XX
PD 09-MAR-2000.
XX
PF 01-SEP-1999; 99MO-US19655.

XX 02-SEP-1998; 98US-0098880.
XX
XX (DIAD-) DIADENUS LLC.
XX
PI Salceda S, Sun Y, Reclon H, Caferkey R;
XX
DR WPI: 2000-256657/22.
DR N-PSDB; AA290478.
XX
PT Diagnosing, staging, monitoring, imaging and treating cancer especially
PT gynecological cancers e.g. breast, ovarian cancer and lung cancer,
PT involves measuring cancer specific gene levels in cells and body fluids
PT -
XX
PS Disclosure; Page 52-54; 58pp; English.
XX
CC The invention relates to detecting, diagnosing metastasis and staging
CC cancer by measuring levels of cancer specific genes (CSG) in cells,
CC tissues or body fluids. Their remission and progression, decreases and
CC increases in CSG levels, is also monitored, by periodic sample analysis.
CC The methods are useful for detecting cancers, especially gynecologic
CC cancers which include ovarian, breast, endometrial and uterine cancer
CC and lung cancer. Antibodies against the CS6s labeled with paramagnetic
CC ions or a radioisotope is useful for imaging cancer and when conjugated
CC with a cytotoxic agent are useful for treating cancer. The present
CC sequence represents a Ovr115 homolog protein, that can be used
CC for the detection of the various cancers.
XX
SQ Sequence 492 AA;

Query Match 98.8%; Score 1303; DB 21; Length 492;
Best Local Similarity 98.3%; Pred. No. 1.8e-115;
Matches 234; Conservative 3; Mismatches 1; Indels 0; Gaps 0;
QY 1 RIVGESALPGAMPMQVSLHVNQVHVCSSITTPBWIYTAHCVEKPLNPNHMTAFAGI 60
DB 255 RIVGESALPGAMPMQVSLHVNQVHVCSSITTPBWIYTAHCVEKPLNPNHMTAFAGI 314
QY 61 LROSFMFAGAGYQVQKVSHPNYSKTKNDIALMKLQKPLTFNDLVKPVCLPDPGMMLQ 120
DB 315 LRSFMTYGAQYQVQKVSHPNYSKTKNDIALMKLQKPLTFNDLVKPVCLPDPGMMLQ 374
QY 121 PEOLCWSIGMGATEEKGKTSVLYNAKVLLETQRCNSRYVDNLITPAMICAGFLQGNV 180
DB 375 PEQLCWSIGMGATEEKGKTSVLYNAKVLLETQRCNSRYVDNLITPAMICAGFLQGNV 434
QY 181 DSCGDSGGLPVTSTKNNIWMILIGDTSMGSCAKATRPVYGVNMFDTMTIRQMRADG 238
DB 435 DSCGDSGGLPVTSTKNNIWMILIGDTSMGSCAKATRPVYGVNMFDTMTIRQMRADG 492

RESULT 7
AA41694
ID AA41694 standard; Protein: 452 AA.
XX
AC AA41694;
XX
DT 07-DEC-1999 (first entry)
XX
DE Human PRO382 protein sequence.
XX
KW Human; PRO; EST; expressed sequence tag; PCR primer; hybridisation;
KW probe; blood coagulation disorder; cancer; cellular adhesion disorder;
KW secreted protein; transmembrane protein.
XX
OS Homo sapiens.
XX
PN WO9946281-A2.
XX
PD 16-SEP-1999.
XX

DT 11-JAN-2001 (first entry)
 XX Tumour associated differentially-expressed gene 12 protein sequence.
 DE Transmembrane serine protease; TADG-12; chromosome 17; vaccination;
 KW tumour associated differentially-expressed gene 12; cytosolic; human;
 KW malignant hyperplasia; cancer; ovary; breast; lung; colon; prostate.
 XX
 OS Homo sapiens.
 XX
 PN MO200052044-A1.
 XX
 PD 08-SEP-2000.
 XX
 PF 02-MAR-2000; 2000MO-US05612.
 XX
 PR 03-MAR-1999; 99US-0261416.
 XX
 PA (UYAR-) UNIV ARKANSAS.
 XX
 PI O'Brien TJ, Underwood LJ;
 XX
 DR WPI: 2000-533263/48.
 DR N-PSDB: AAA93842.
 XX
 PT DNA fragment encoding tumor associated differentially-expressed gene 12
 PT protein used for diagnosing and treating malignant hyperplasia and
 PT cancers including ovarian cancer -
 XX
 PS Claim 3; Figure 4; 118bp; English.
 XX
 CC This invention relates to a novel transmembrane serine protease called
 CC tumour associated differentially-expressed gene 12 (TADG-12). TADG-12 is
 CC located on chromosome 17. Sequences AAA93842-A93845 and AAB32246-B32249
 CC represent human TADG-12 cDNA and their corresponding protein sequences.
 CC A splice variant of TADG-12 (TADG-12V) leads to a truncated protein
 CC product. TADG-12 is overexpressed in ovarian carcinomas. TADG-12
 CC exhibits cytosolic activity, and can be used in vaccines and in gene
 CC therapy. TADG-12 nucleotide and protein sequences are used in the
 CC diagnosis of malignant hyperplasia and cancers of the ovary, breast,
 CC lung, colon, prostate and other cancers where TADG-12 is overexpressed.
 CC TADG-12 is particularly used as tumor marker for early disease
 CC diagnosis. TADG12 proteins or fragments can be used to vaccinate an
 CC individual with cancer, suspected of having a cancer or at risk of
 CC getting cancer. Sequences AAA93846-A93853 represent PCR primers used for
 CC amplifying the TADG-12 cDNA sequence, and in the quantitative analysis of
 CC TADG-12 mRNA. AAB32250 represents a peptide fragment of TADG-12, used to
 CC create anti-TADG-12 antibodies. Sequences AAB32251-B32369 represent
 CC TADG-12 peptides which target HLA, and may be used in a vaccine or for
 CC immune stimulation.
 CC
 XX
 SQ Sequence 454 AA;
 Query Match 49.7%; Score 655; DB 21; Length 454;
 Best Local Similarity 52.9%; Pred. No. 4.5e-54;
 Matches 126; Conservative 30; Mismatches 78; Indels 4; Gaps 3;

RESULT 9
 AAB44250
 ID AAB44250 standard; Protein: 453 AA.
 XX
 AC AAB44250;
 XX
 DT 08-FEB-2001 (first entry)
 XX
 DE Human PRO382 (UNQ323) protein sequence SEQ ID NO:69.
 XX
 KW Human; secreted protein; transmembrane protein; PRO; EST; cytosolic;
 KW expressed sequence tag; detection; cancer.
 XX
 OS Homo sapiens.
 XX
 PN MO200053756-A2.
 XX
 PD 14-SEP-2000.
 XX
 PF 18-FEB-2000; 2000MO-US04341.
 XX
 PR 08-MAR-1999; 99MO-US05028.
 PR 12-MAR-1999; 99US-0123957.
 PR 29-MAR-1999; 99US-0126773.
 PR 21-APR-1999; 99US-0130232.
 PR 28-APR-1999; 99US-0131445.
 PR 14-MAY-1999; 99US-0134287.
 PR 23-JUN-1999; 99US-0141037.
 PR 26-JUL-1999; 99US-0145698.
 PR 29-OCT-1999; 99US-0162506.
 PR 30-NOV-1999; 99MO-US28513.
 PR 02-DEC-1999; 99MO-US28551.
 PR 16-DEC-1999; 99MO-US28565.
 PR 30-DEC-1999; 99MO-US30095.
 PR 30-DEC-1999; 99MO-US31243.
 PR 05-JAN-2000; 2000MO-US01219.
 PR 06-JAN-2000; 2000MO-US00277.
 PR 06-JAN-2000; 2000MO-US00376.
 XX
 PA (GETH) GENENTECH INC.
 XX
 PI Ashkenazi AJ, Baker KP, Botstein D, Desnoyers L, Eaton DL,
 PI Ferrara N, Filvaroff E, Fong S, Gao W, Gerber H, Gerritsen ME,
 PI Goddard A, Godowski PJ, Grimaldi CJ, Gurney AL, Hillan KJ,
 PI Kljavin IJ, Kuo SS, Napier MA, Pan J, Paoni NF, Roy MA,
 PI Shelton DL, Stewart TA, Tumas D, Williams PM, Wood WI;
 XX
 DR WPI: 2000-611443/58.
 DR N-PSDB: AAC78475.
 XX
 PT Novel PRO polypeptides and polynucleotides used in detection methods,
 PT to target bioactive molecules to specific cells, and to modulate
 PT cellular activities -
 XX
 PS Claim 12; Fig 28; 636bp; English.
 XX
 CC AAC78458 to AAC78599 represent polynucleotide and EST (expressed
 CC sequence tag) sequences which encode secreted or transmembrane PRO
 CC polypeptides. The PRO polynucleotides and polypeptides have cytosolic
 CC activity. The polynucleotides and polypeptides can be used for detecting
 CC the presence of PRO polypeptides in samples, for linking bioactive
 CC molecules to cells and for modulating biological activities of cells.
 CC using the polypeptides for specific targeting. The polypeptide targeting
 CC can be used to kill the target cells, e.g. for the treatment of cancers.
 CC The polypeptide pairs provide specific targeting of bioactive molecules
 CC to cells. AAC78600 to AAC78987 represent PCR primers and probes used in
 CC the isolation of the PRO polynucleotide sequences.
 XX
 SQ Sequence 453 AA;

RESULT	10
AA043572	
ID	AA043572 standard; Protein: 248 AA.
XX	
AC	AA043572;
XX	
DT	08-FEB-2001 (first entry)
XX	
DE	Human cancer associated protein sequence SEQ ID NO:1017.

KW Human; cancer associated gene; cancer antigen; detection; cancer;
 KW diagnosis; cytotoxic; proliferative; antineoplastic; immunomodulator;
 KW antidiabetic; antihistaminic; antirheumatic; antiarthritic; antiviral;
 KW antiinflammatory; antithyroid; antiallergic; antibacterial; cardiac;
 KW dermatological; neuroprotective; thrombolytic; coagulant; neutropic;
 KW vasotropic; antipsoriatic; antiangiogenic; gene therapy; inflammation;
 KW immune disorder; haematopoietic cell disorder; autoimmune disorder;
 KW allergic reaction; graft versus host disease; organ rejection;
 KW haemostatic; thrombolytic; cardiovascular disorder; infection;
 KW neurological disease; drug screening.
 OS Homo sapiens.
 XX
 PN WO200055350-A1.
 XX
 PD 21-SEP-2000.
 XX
 PF 08-MAR-2000; 2000WO-US05882.
 XX
 PR 12-MAR-1999; 99US-0124270.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Rosen CA, Ruben SM;.
 XX
 DR WPI: 2000-587533/55.
 DR N-PSDB; AAC77781.
 XX
 PT Novel isolated nucleic acids comprising sequences encoding peptides
 XX useful for treating or diagnosing e.g. cancer -
 XX
 Claim 11; Page 1598-1599; 2352pp; English.

CC AABC7667.00 AABC7844.00 encode the human cancer associated proteins given
CC in AABC3368.00 to AABC44239.00. The proteins can have activities based on the
CC tissues and cells the genes are expressed in. Example of activities
CC include: cytostatic; proliferative; vulnerrary; immunomodulator;
CC antidiabetic; antiasthmatic; antirheumatic; antiarthritic;
CC antinflammatory; antithyroid; antiallergic; antibacterial; antiviral;
CC dermatologic; neuroprotective; cardiac; thrombolytic; coagulant;
CC neurotropic; vasotropic; antiprolastic and antiangiogenic. CC

Query Match	48.9%	Score	645.5	DB	21	Length	248
Best Local Similarity	52.7%	Pred. No.	1.7e-53				
Matches 126; Conservative	30;	Mismatches	78;			Indels	5;
						Gaps	4;

```

QY 1 RIVGESALPGAMWPOVSLHVNQNVCCSGITTPREWITAAICVEKRLPNHWHAFAGI 60
Db 10 rlvvgmsmlsrgwqbasqfiggyhlccgsvvltplwiltaaicv-ydllylprksvdiqvl 68
QY 61 LROSMFYGCA-GYQVEKVISHPNDKSTNNNDIALMKLOKPLTFENDLVKPYCLPPRGML 119
Db 69 v---slldnpapshlyekilyhskypkrlgndialmklaagpltfemcgvpcslpseenf 126
QY 120 OPEOLCWTISGMCATEE-KGKTSEVNAAKVLDETORCNSRYVDNLITTPAMICAFLOG 178
Db 127 pdkgkcwtswgatededgagdaasvlnhaavplnsnkicmhrcdyvggslspmlcagyltg 186
QY 179 NVDSQCGDSGGLVYNSKNNIMWLIGDTISKSGSCAAKAPRGVYGNVWTTDVIYRMRAD 237
Db 187 gvdscgsgdsagglvcqerrlwlklyvatsfigldeavmnpvyvtyrltsldwlshegmed 245

```

AC AAY72093;

DT 28-MAR-2001 (first entry)

DE Human serine protease #4 encoded by clone HCHAK72.

KM Human; serine protease; osteopathic; immunosuppressive; antiallergic;
 KM antineoplastic; cytostatic; cardiatic; neuroprotective; neurologic;
 KM antiinflammatory; vulnery; ophthalmologic; antibacterial; antiviral;
 KM antifungal; antiparasitic; gene therapy; diagnosis; prevention; glaucoma;
 KM treatment; bone formation disorder; osteoporosis; arthritis; cancer;
 KM connective tissue disorder; autoimmune disorder; wound healing; asthma;
 KM systemic lupus erythematosus; male reproductive system disorder;
 KM testicular cancer; digestion and food absorption disorder; arhythmia;
 KM Crohn's disease; neurodegenerative disease; Alzheimer's disease; allergy;
 KM behavioral disorder; Tourette's syndrome; acute myelogenous leukaemia;
 KM cardiovascular disorder; ocular disorder; drug screening.

Homo sapiens.

PN W0200068247-A2.

PD 16-NOV-2000

PF 05-MAY-2000; 2000WO-US12207.

PR 07-MAY-1999; 99US-0133239.

PR 03-AUG-1999; 99US-0147005.

PR 01-NOV-1999; 99US-0162979

XX (HUMA-) HUMAN GENOME SCI INC.
 PA Ruben SM, Shi Y, Young PE, Ni J;
 PI WPI: 2000-679799/66.
 XX N-PSDB: AAD02323.
 DR
 XX New nucleic acid molecules encoding human serine protease polypeptides,
 PT useful for diagnosis, prevention and/or treatment of disorders e.g.
 PT osteoporosis, lupus erythematosus and Alzheimer's -
 PS
 XX Claim 12: Page 280-281; 289pp: English.
 XX
 XX The present sequence is human serine protease #4 from clone
 CC HCHAK72 (ATCC Deposit No: PTA27).
 CC The invention relates to human serine proteases and their cDNA clones.
 CC It is used in methods for the diagnosis, prevention and treatment of
 CC various disorders related to serine protease such as bone formation
 CC disorders (osteoporosis), connective tissue disorders (arthritis),
 CC autoimmune disorders (systemic lupus erythematosus), wound healing, male
 CC reproductive system disorders (testicular cancer), digestion and food
 CC absorption disorders (Crohn's disease), neurodegenerative diseases
 CC (Alzheimer's disease), behavioral disorders (Tourette's syndrome),
 CC proliferative and cancerous conditions (acute myelogenous leukemia),
 CC allergic reactions (asthma), cardiovascular disorders (arrhythmia),
 CC ocular disorders (glaucoma) and infectious diseases caused by bacteria,
 CC viruses, fungi or parasites. It is also useful for screening therapeutic
 CC compounds. Serine proteases are used as immunological probes or
 CC polymorphic markers for the identification of chromosomes, cells and
 CC tissues in biological samples, identification of male contraceptive
 CC agents, delivery of compositions to targeted cells expressing a
 CC receptor for serine protease, hybridisation probes and molecular weight
 CC markers. Serine protease nucleic acids are also useful in gene therapy.
 CC Note: The present sequence shown in page 280-281 of sequence listing has
 CC been assigned SEQ ID NO: 14. But the sequence, human serine protease
 CC epidermal growth factor (EGF)-like domain (AAV2114) shown in
 CC page 12 is also referred as SEQ ID NO: 14.
 CC
 XX
 XX Sequence 327 AA;
 SQ
 Query Match 48.9%; Score 645.5; DB 21; Length 327;
 Best Local Similarity 52.7%; Pred. No. 2.4e-53;
 Matches 126; Conservative 30; Mismatches 78; Indels 5; Gaps 4;
 QY 1 RIVGESALPGAMPQVSLHVNVAHVCSSITTPMIYTAACVCEKPLNNPMHMTAFAGI 60
 Db 89 RIVGSMALLSGWPQGAALGFQGYHLGGSVLTPLWILTAACV-ydlylpkswclqygl 147
 QY 61 LRQSFMEYGA-GYQVEKVISHPNYSKTKNDIALMLQKPLTFNDLVKPYCLPNPGMML 119
 Db 148 V--allndpaphplvlekiyhsykrkrlgnldaimklagpltfnemlqpvclpseenf 205
 QY 120 QPEQLCHISGKATEE-KGKTSEVLNAKVLIIETORNSRKYVNDLITPAICGFLQG 178
 Db 206 pdgkvcwsgagatgedagdasprlnhaavplsnkicnhdvvg9glispsmlcsgyltg 265
 QY 179 NVDSQGDGSGPLVYTSKNNIMWLIGDTSMSGSCAKARPGYGVNVMVPTDIYRMRAD 237
 Db 266 gydsqgsgspplvcegrllwklvgatstfgcgaevnkpvytlrtstfldwhegmard 324
 RESULT 12
 AAB08912
 ID AAB08912 standard; Protein: 414 AA.
 AC AAB08912;
 XX
 XX 30-AUG-2000 (first entry)
 XX Human secreted protein sequence encoded by gene 22 SPO ID NO:69.
 DE

KW Human; secreted protein; cytosolic; anti-proliferative; vulnery;
 KW immunosuppressive; antibacterial; diagnosis; immune system; chemotaxis;
 KW hyperproliferative disorder; infectious disease; tissue regeneration;
 KW screening; food additive; preservative; wound healing;
 KW hyper-vascular disease.
 OS Homo sapiens.
 PN WO200017222-A1.
 XX
 XX 30-MAR-2000.
 PD
 XX 22-SEP-1999; 99WO-US22012.
 PF
 XX 23-SEP-1998; 98US-0101546.
 PR 02-OCT-1998; 98US-0102895.
 XX
 XX (HUMA-) HUMAN GENOME SCI INC.
 PA Ruben SM, Rosen CA, Duan RD, Shi Y, Lafleur DW, Young PE, Ni J;
 PI Komatsoulis G, Endress GA, Soppet DR;
 XX WPI: 2000-283538/24.
 DR N-PSDB: AAA39073.
 PS
 XX Claim 11: Page 367-368; 416pp: English.
 CC The polynucleotide sequences given in AAA39052 to AAA39088 encode the
 CC human secreted proteins given in AAB08891 to AAB08984. The human secreted
 CC proteins can have activities based on the tissues and cells they are
 CC expressed in. Examples of the activities are: cytostatic;
 CC anti-proliferative; immunosuppressive; antibacterial; and vulnery. The
 CC secreted proteins and their related polynucleotide sequences are useful
 CC for diagnostic and therapeutic methods useful for diagnosing and treating
 CC disorders related to the secreted proteins. The proteins, and
 CC polynucleotide sequences may be useful for treating disorders of the
 CC immune system, hyperproliferative disorders, infectious disease,
 CC regeneration of tissues, for chemotaxis and for screening molecules that
 CC bind to the proteins. The proteins or polynucleotide sequences may be
 CC used as food additives or preservatives. To increase or decrease storage
 CC capabilities, fat content, lipid, protein, carbohydrate, vitamins,
 CC minerals, co-factors or other nutritional components. Atonists or
 CC antagonists of the proteins may be used to prevent scar tissue growth
 CC during wound healing, and hyper-vascular diseases. AAA39043 to AAA39051
 CC and AAB08890 are sequences used in the exemplification of the present
 CC invention.
 CC
 XX
 XX Sequence 414 AA;
 SQ
 Query Match 42.8%; Score 564; DB 21; Length 414;
 Best Local Similarity 45.3%; Pred. No. 1.7e-45;
 Matches 106; Conservative 31; Mismatches 94; Indels 2; Gaps 2;
 QY 1 RIVGESALPGAMPQVSLHVNVAHVCSSITTPMIYTAACVCEK-PLNNPMHMTAFAG 59
 Db 173 RIVGSGVAPRGWPQGAVALGFHTCGSVLAPRVVLTAAHCHMSFIALISRYRHAG 232
 QY 60 ILRQSFMEYGAQYQVEKVISHPNYSKTKNDIALMLQKPLTFNDLVKPYCLPNPGMML 119
 Db 233 lvshavprhgalverllrplysaqnhdvvalrlrtalntsdvavclpakeghf 292
 QY 120 QPEQLCHISGKATEEKGK-TSEVLNAKVLIIETORNSRKYVNDLITPAICGFLQG 178
 Db 293 pkgsrvcwsgwghthpshlytsdmlqdtvplfstqjcnsscvcysgaltpmlcaayldg 352
 QY 179 NVDSQGDGSGPLVYTSKNNIMWLIGDTSMSGSCAKARPGYGVNVMVPTDIY 231
 Db 353 radacqsgspplvcepdgtlwrivgvswwrgcdephpyyavkaevfllwln 405

RESULT 13
ID AAB08950 standard; Protein; 480 AA.
XX
AC AAB08950;
XX
DT 30-AUG-2000 (first entry)
XX
DE Human secreted protein sequence encoded by gene 22 SEQ ID NO:107.
XX
KW Human; secreted protein; cytosolic; anti-proliferative; vulnery;
KW immunosuppressive; antibacterial; diagnosis; immune system; chemotaxis;
KW hyperproliferative disorder; infectious disease; tissue regeneration;
KW screening; food additive; preservative; wound healing;
KW hyper-vascular disease.
XX
XX Homo sapiens.
OS
PN MO200017222-A1.
XX
PD 30-MAR-2000.
XX
PF 22-SEP-1999; 99WO-US22012.
XX
PR 23-SEP-1998; 98US-0101546.
PR 02-OCT-1998; 98US-0102895.
XX
PA (HDMA-) HUMAN GENOME SCI INC.
XX
PI Ruben SM, Rosen CA, Duan RD, Shi Y, Lafleur DW, Young PE, Ni J;
PI Komatsoulis G, Endress GA, Soppet DR;
XX WPI: 2000-283538/24.
XX
DR Human secreted proteins and coding sequences useful in diagnostic and
XX therapeutic methods for disorders such as immune system or
XX proliferative disorders, related to the proteins
XX
PS Disclosure: Page 65; 416pp; English.
XX
CC The polynucleotide sequences given in AAA39052 to AAA39088 encode the
CC human secreted proteins given in AAB08891 to AAB08984. The human secreted
CC proteins can have activities based on the tissues and cells they are
CC expressed in. Examples of the activities are: cytostatic;
CC anti-proliferative; immunosuppressive; antibacterial; and vulnery. The
CC secreted proteins and their related polynucleotide sequences are useful
CC for diagnostic and therapeutic methods useful for diagnosing and treating
CC disorders related to the secreted proteins. The proteins, and
CC polynucleotide sequences may be useful for treating disorders of the
CC immune system, hyperproliferative disorders, infectious disease,
CC regeneration of tissues, for chemotaxis and for screening molecules that
CC bind to the proteins. The proteins or polynucleotide sequences may be
CC used as food additives or preservatives, to increase or decrease storage
CC capabilities, fat content, lipid, protein, carbohydrate, vitamins,
CC minerals, co-factors or other nutritional components. Agonists or
CC antagonists of the proteins may be used to prevent scar tissue growth
CC during wound healing, and hyper-vascular diseases. AAA39043 to AAA39051
CC and AAB08890 are sequences used in the exemplification of the present
CC invention.
XX
SQ Sequence 480 AA;

Query Match 42.8%; Score 564; DB 21; Length 480;
Best Local Similarity 45.5%; Pred. No. 2e-45;
Matches 106; Conservative 31; Mismatches 94; Indels 2; Gaps 2;

QY 1 RIVGGSALPGAMPQVSLHVNQVHCGSIIITPEWIVTAHAKVER-PLNPNMHTAFAG 59
DB 240 RIVGGSVAPGTPWQVAVSLGFRHCGSVLPRVWVLAHCHMSFRLARLISWTVHAG 299

QY 60 ILRQSEFMFGAGYQVEKVISHPNYSKTRKNNDIALKLOKPLTFNDLPVCLPNPQML 119
DB 300 lvsasvrbphgdlverilphlypsagmndyavallrlqtafnfstbtavclpakeqnf 359
QY 120 QPEQLCWISGWCATEEKKR-TSEVLNAKAVLLIETORCSRYYDNLITPAMICAFLOG 178
DB 360 pkgrcwsvsgwghnphshyscdmlqdtvplfstqlcnsscvsygalprmlcagylgd 419
QY 179 NVDSGCGDSCGLVTSKNNIMWLIQDTSMGSCGCAKVRGCVGNVVFDMY 231
DB 420 radacqdsqgplvcpgdltwrlivgvsrgcaepnhpyakvaefldwln 472

RESULT 14
ID AAY72108 standard; Protein; 296 AA.
XX
AC AAY72108;
XX
DT 28-MAR-2001 (first entry)
XX
DE Human serine protease #3 fragment from HATEB38 cDNA clone.
XX
KW Human; serine protease; osteopathic; immunosuppressive; anti-allergic;
KW anti-inflammatory; cytostatic; cardiant; neuroprotective; nocrotic;
KW neuroleptic; vulnery; ophthalmological; antibacterial; antiviral;
KW antifungal; antiparasitic; gene therapy; diagnosis; prevention; gliacoma;
KW treatment; bone formation disorder; osteoporosis; arthritis; cancer;
KW connective tissue disorder; autoimmune disorder; wound healing; asthma;
KW systemic lupus erythematosus; male reproductive system disorder;
KW testicular cancer; digestion and food absorption disorder; arrhythmia;
KW Crohn's disease; neurodegenerative disease; Alzheimer's disease; allergy;
KW behavioral disorder; Tourette's syndrome; acute myelogenous leukaemia;
KW cardiovascular disorder; ocular disorder; drug screening.
XX
XX Homo sapiens.
OS
XX
PN MO200068247-A2.
XX
PD 16-NOV-2000.
XX
PF 05-MAY-2000; 2000WO-US12207.
XX
PR 07-MAY-1999; 99US-0133239.
PR 20-MAY-1999; 99US-0135163.
PR 03-AUG-1998; 99US-0147005.
PR 09-SEP-1999; 99US-0152935.
PR 01-NOV-1999; 99US-0162979.
XX
PA (HDMA-) HUMAN GENOME SCI INC.
XX
PI Ruben SM, Shi Y, Young PE, Ni J;
XX WPI: 2000-679799/66.
XX N-PSDB: AAD02322.
XX
PT New nucleic acid molecules encoding human serine protease polypeptides,
PT useful for diagnosis, prevention and/or treatment of disorders e.g.
PT osteoporosis, lupus erythematosus and Alzheimer's -
XX
PS Disclosure: Page 286-287; 289pp; English.

XX The present sequence is human serine protease #3 fragment from clone
CC HATEB38 (ATCC Deposit No. PTA27).
CC The invention relates to human serine proteases and their cDNA clones.
CC It is used in methods for the diagnosis, prevention and treatment of
CC various disorders related to serine protease such as bone formation
CC disorders (osteoporosis), connective tissue disorders (arthritis),
CC autoimmune disorders (systemic lupus erythematosus), wound healing, male
CC reproductive system disorders (testicular cancer), digestion and food
CC absorption disorders (Crohn's disease), neurodegenerative diseases
CC (Alzheimer's disease), behavioral disorders (Tourette's syndrome),
CC proliferative and cancerous conditions (acute myelogenous leukaemia),

CC allergic reactions (asthma), cardiovascular disorders (arrhythmia),
CC ocular disorders (glaucoma) and infectious diseases caused by bacteria,
CC viruses, fungi or parasites. It is also useful for screening therapeutic
CC compounds. Serine proteases are used as immunological probes or
CC polymorphic markers for the identification of chromosomes, cells and
CC tissues in biological samples, identification of male contraceptive
CC agents, delivery of compositions to targeted cells expressing a
CC receptor for serine protease, hybridisation probes and molecular weight
CC markers. Serine protease nucleic acids are also useful in gene therapy.
CC Note: The present sequence shown in page 286-287 of sequence listing has
CC been assigned SEQ ID NO: 28. But the sequence, peptide fragment #25
CC related to human serine protease (AA72128) shown in page 22 is also
CC referred as SEQ ID NO: 28.

CC Sequence 296 AA:

Query Match 42.3%; Score 558; DB 21; Length 296;
Best Local Similarity 45.6%; Pred. No. 4.2e-45;
Matches 109; Conservative 34; Mismatches 86; Indels 10; Gaps 4;

OY 1 RIVGESALPGAMPQVSLHVNHVCGSITTPMITYAAHCV---EKPLNPMHMTA 56
DB 4 RIVGALASDskwpqvsahfgthlhcggclldagwvltahcfcvtrckvleg---kvv 60
OY 57 FAGILRSFMFYGAGYQVEKVISHPNDSKTKNNDIAMKLQKPLTFNDLVKPCLPNG 116
DB 61 YAGTGNLHQLPEAA--sfaellinsnytdceddydlamrlskphtlsahhpacplmgh 118
OY 117 MMLQPEQLCWSIGMGATEE-KGKTSEVLNAAKVLLEIQRCSRRVYDNLITPAMICAGF 175
DB 119 qtfslncwclwlgfgrtkretddktsplfrevgnltdfkckndylyvdsyltprmmcagd 178
OY 176 LOGNVDSOGSGGGLVYTSKNNIMWLIGDTSNGSCAAKAPRGVYGNVWFTDWTYROM 234
DB 179 LRgrdsccgsgsgplvcegnmrwylagvtswgicgqgrnkpytkvlelvpwlyskm 237

RESULT 15
AA72092
ID AA72092 standard; Protein: 372 AA.

XX AA72092;
XX 28-MAR-2001 (first entry)
XX Human serine protease #3 encoded by clone HATEE38.
XX
KW Human; serine protease; osteopathic; immunosuppressive; antiallergic;
KW antiinflammatory; cytostatic; cardiac; neuroprotective; nocitropic;
KW neuroleptic; vulnery; ophthalmological; antibacterial; antiviral;
KW antifungal; antiparasitic; gene therapy; diagnosis; prevention; glaucoma;
KW treatment; bone formation disorder; osteoporosis; arthritis; cancer;
KW connective tissue disorder; autoimmune disorder; wound healing; asthma;
KW systemic lupus erythematosus; male reproductive system disorder;
KW testicular cancer; digestion and food absorption disorder; arrhythmia;
KW Crohn's disease; neurodegenerative disease; Alzheimer's disease; allergy;
KW behavioral disorder; Tourette's syndrome; acute myelogenous leukaemia;
KW cardiovascular disorder; ocular disorder; drug screening.

XX OS Homo sapiens.
XX XX
XX PN MO200068247-A2.
XX PD 16-NOV-2000.
XX PE 05-MAY-2000; 2000MO-US12207.
XX PR 07-MAY-1999; 99US-0133239.
XX PR 20-MAY-1999; 99US-0135163.
XX PR 03-AUG-1999; 99US-0147005.
XX PR 09-SEP-1999; 99US-0152935.
XX PR 01-NOV-1999; 99US-0162979.

XX (HUMA-) HUMAN GENOME SCI INC.
PA
XX Ruben SM, Shi Y, Young PE, Ni J;
PI WPI: 2000-679799/66.
XX N-PSDB: AAD02322.
DR

PT New nucleic acid molecules encoding human serine protease polypeptides,
PT useful for diagnosis, prevention and/or treatment of disorders e.g.
PT osteoporosis, lupus erythematosus and Alzheimer's -
XX
XX Claim 12; Page 278-280; 289pp; English.

XX The present sequence is human serine protease #3 from clone
XX HATEE38 (ATCC Deposit No: PTA27).

XX The invention relates to human serine proteases and their cDNA clones.
XX It is used in methods for the diagnosis, prevention and treatment of
XX various disorders related to serine protease such as bone formation
XX disorders (osteoporosis), connective tissue disorders (arthritis),
XX autoimmune disorders (systemic lupus erythematosus), wound healing, male
XX reproductive system disorders (testicular cancer), digestion and food
XX absorption disorders (Crohn's disease), neurodegenerative diseases
XX (Alzheimer's disease), behavioral disorders (Tourette's syndrome),
XX proliferative and cancerous conditions (acute myelogenous leukaemia),
XX allergic reactions (asthma), cardiovascular disorders (arrhythmia),
XX ocular disorders (glaucoma) and infectious diseases caused by bacteria,
XX viruses, fungi or parasites. It is also useful for screening therapeutic
XX compounds. Serine proteases are used as immunological probes or
XX polymorphic markers for the identification of chromosomes, cells and
XX tissues in biological samples, identification of male contraceptive
XX agents, delivery of compositions to targeted cells expressing a
XX receptor for serine protease, hybridisation probes and molecular weight
XX markers. Serine protease nucleic acids are also useful in gene therapy.

XX Sequence 372 AA:

Query Match 42.3%; Score 558; DB 21; Length 372;
Best Local Similarity 45.6%; Pred. No. 5.6e-45;
Matches 109; Conservative 34; Mismatches 86; Indels 10; Gaps 4;

OY 1 RIVGESALPGAMPQVSLHVNHVCGSITTPMITYAAHCV---EKPLNPMHMTA 56
DB 4 RIVGALASDskwpqvsahfgthlhcggclldagwvltahcfcvtrckvleg---kvv 60
OY 57 FAGILRSFMFYGAGYQVEKVISHPNDSKTKNNDIAMKLQKPLTFNDLVKPCLPNG 116
DB 61 YAGTGNLHQLPEAA--sfaellinsnytdceddydlamrlskphtlsahhpacplmgh 118
OY 117 MMLQPEQLCWSIGMGATEE-KGKTSEVLNAAKVLLEIQRCSRRVYDNLITPAMICAGF 175
DB 119 qtfslncwclwlgfgrtkretddktsplfrevgnltdfkckndylyvdsyltprmmcagd 178
OY 176 LOGNVDSOGSGGGLVYTSKNNIMWLIGDTSNGSCAAKAPRGVYGNVWFTDWTYROM 234
DB 179 LRgrdsccgsgsgplvcegnmrwylagvtswgicgqgrnkpytkvlelvpwlyskm 237

Search completed: September 26, 2001, 09:22:50
Job time: 105 sec

THIS PAGE BLANK (USPTO)

Gencore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 26, 2001, 09:21:42 : Search time 25.88 Seconds

(without alignments)
700.523 Million cell updates/sec

Title: US-09-615-285-2_COPY_255_492

Perfect score: 1319

Sequence: 1 RYGGESALPGAMPQVSLH.....VYGNVWFDTQWYRQNRADG 238

Scoring table: BLOSUM62

Gapop 10.0, Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : PIR-68:*

1: p1r1:*

2: p1r2:*

3: p1r3:*

4: p1r4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	558.5	42.3	638	1	KOHUP
2	543	41.2	638	1	KORPL
3	530	40.2	638	1	KOMSPL
4	529	40.1	625	1	KFHUI
5	528.5	40.1	417	1	S00845
6	523.5	39.7	1035	1	A43090
7	513.5	38.9	1019	1	A56318
8	502.5	38.1	812	1	PLMS
9	499.5	37.9	416	1	S33777
10	499	37.8	1524	2	T30337
11	492	37.3	421	1	S11674
12	488.5	37.0	1034	1	A53663
13	481.5	36.5	415	1	A34170
14	476.5	36.1	436	2	UX0172
15	466.5	35.4	431	2	S47538
16	464.5	35.2	421	2	S29599
17	464	35.2	418	2	A37344
18	464	35.2	790	1	PLPG
19	460.5	34.9	437	2	S18407
20	458	34.7	810	2	B30848
21	457.5	34.7	810	2	I46260
22	454	34.4	460	2	B61545
23	451.5	34.2	812	1	PLBO
24	450.5	34.2	267	2	S40006
25	450	34.1	270	2	S56160
26	449	34.0	275	2	S40005
27	449	34.0	343	1	A57014
28	447	33.9	276	2	A38654
29	447	33.9	277	2	S35340

ALIGNMENTS

30	443	33.6	455	2	A61545	plasmin (EC 3.4.21
31	442.5	33.5	266	2	S54146	trypsin (EC 3.4.21
32	442.5	33.5	4548	1	S00657	apoptein(a) (EC
33	442	33.5	810	1	PLMU	plasmin (EC 3.4.21
34	441	33.4	1113	2	JE0315	low-density lipopr
35	440	33.4	274	2	JC4171	trypsin (EC 3.4.2
36	439	33.3	273	2	A47246	trypsin (EC 3.4.2
37	438	33.2	274	2	S35339	trypsin (EC 3.4.21
38	437	33.1	275	2	S40007	trypsin (EC 3.4.21
39	433.5	32.9	420	2	A55283	acrosin (EC 3.4.21
40	432	32.8	242	2	S49489	trypsin (EC 3.4.21
41	432	32.8	275	2	C35863	trypsin (EC 3.4.2
42	430	32.6	244	2	S72219	chymotrypsin B - A
43	430	32.6	275	2	A35863	trypsin (EC 3.4.2
44	429	32.5	263	1	KYRTR	chymotrypsin (EC 3
45	429	32.5	275	2	B35863	trypsin (EC 3.4.2

RESULT 1

KOHUP
plasmin kallikrein (EC 3.4.21.34) precursor - human
M:Alternate names: kininogenin; plasma prekallikrein
C:Species: Homo sapiens (man)
C:Date: 13-Aug-1986 #sequence.revision 13-Aug-1986 #text.change 18-Jun-1999
C:Accession: A00921; A37939
R:Chung, D.W.; Fujikawa, K.; McMullen, B.A.; Davie, E.W.
Biochemistry 25, 2410-2417, 1986
A:Title: Human plasma prekallikrein, a zymogen to a serine protease that contains fou
A:Reference number: A00921; MUID:86243359
A:Accession: A00921
A:Molecule type: mRNA
A:Residues: 1-638 <CHD>
A:Cross-references: GB:M13143; NID:9190262; PIDN:AAA60153.1; PID:9190263
M:McMullen, B.A.; Fujikawa, K.; Davie, E.W.
Biochemistry 30, 2050-2056, 1991
A:Title: Location of the disulfide bonds in human plasma prekallikrein: the presence
A:Reference number: A37939; MUID:91152016
A:Accession: A37939
A:Molecule type: protein
A:Residues: 20-27;40-46; 'X',48; 'H',50; 'X',52-70; 'H',75-76; 'X',78-80;103-113;131-140;1
;260-283; 'X',285;287-291; 'X',293-295;314-317; 'X',319-320;321-324; 'X',329-333;334-339,
525;538-551;562; 'X',564-567;573; 'X',575-576;578-583; 'X',585;592-604 <MCK>
C:Comment: This protein, synthesized in the liver, circulates as a noncovalent comp
C:Comment: The zymogen is activated by factor XIIa, which cleaves the molecule into a
are linked by one or more disulfide bonds.
C:Comment: The enzyme cleaves Lys-Arg and Arg-Ser bonds. It activates, in a reciproca
inogen and may also play a role in the renin-angiotensin system by converting proren
C:Genetics: GDB:KLR3
A:Gene: GDB:KLR3
A:Cross-references: GDB:127575; OMIM:229000
A:Map position: 4q35-4q35
C:Keywords: blood coagulation; duplication; fibrinolysis; glycoprotein; hydrolase; in
F:1-19/Domain: signal sequence #status predicted <SIC>
F:20-638/Product: plasma kallikrein #status predicted <MAT>
F:20-390/Domain: plasma kallikrein heavy chain #status predicted <HCH>
F:20-109/Domain: apple repeat <AP1>
F:110-199/Domain: apple repeat <AP2>
F:200-289/Domain: apple repeat <AP3>
F:291-380/Domain: apple repeat <AP4>
F:391-638/Domain: plasma kallikrein light chain #status predicted <LCH>
F:391-621/Domain: trypsin homology <TR>
F:21-104;47-77;51-57;111-194;137-166;141-147;201-284;227-256;231-237;292-375;322-328,
F:127;104;396;453;494/Binding site: carboxylate (Asn) (covalent) #status experimenta
F:318-347;340-345/Disulfide bonds: #status predicted
F:390-391/Cleavage site: Arg-Ile (coagulation factor XIIa) #status predicted
F:434;483;578/Active site: His, Asp, Ser #status predicted

Query Match

42.3%; Score 558.5; DB 1; Length 638;

[illegible]

RESULT 2

KORTPL

plasma kallikrein (EC 3.4.21.34) precursor - rat

N:Alternate names: Fletcher factor; kininogenin; serum kallikrein

C:Species: Rattus norvegicus (Norway rat)

C:Date: 30-Sep-1992 #sequence_revision 30-Sep-1992 #text_change 18-Jun-1999

C:Accession: A39180; A33320; S06851; I53041; S06852

R:Beaubien, G.; Rosinski-Chupin, I.; Mattei, M.G.; Mlikay, M.; Chretien, M.; Seldah, N.G.

A:Title: Gene structure and chromosomal localization of plasma kallikrein.

A:Accession: A39180

A:Molecule type: DNA

A:Residues: 1-638

A:Cross-references: GB:J05315

A:Note: The authors translated the codon GAG for residue 81 as Gln

R:Seldah, N.G.; Ladenheim, R.; Mlikay, M.; Hamelin, J.; Lutfalla, G.; Rougeon, F.; Lazur

DNA 8, 563-574, 1989

A:Title: The cDNA structure of rat plasma kallikrein.

A:Reference number: A33320; MUID:90091743

A:Accession: A33320

A>Status: not compared with conceptual translation

A:Residues: 1-638 <SR>

A:Molecule type: mRNA

A:Cross-references: GB:M30282; NID:g205010; PIDN:AAA4163.1; PID:g205011

A:Note: part of this sequence, including the amino ends of both the heavy and light chain

R:Paguin, J.; Benjannet, S.; Sawyer, N.; Lazure, C.; Chretien, M.; Seldah, N.G.

Biochim. Biophys. Acta 999, 103-110, 1989

A:Title: Rat plasma kallikrein: purification, NH(2)-terminal sequencing and development

A:Reference number: S06851; MUID:90089457

A:Accession: S06851

A:Molecule type: Protein

A:Residues: 20-45;391-413 <PAQ>

R:Seldah, N.G.; Ladenheim, R.; Mlikay, M.; Hamelin, J.; Lutfalla, G.; Rougeon, R.; Lazur

DNA Cell Biol. 8, 563-574, 1989

A:Title: The cDNA structure of rat plasma kallikrein.

A:Reference number: I53041

A:Accession: I53041

A>Status: translated from GB/EMBL/DDBJ

A:Molecule type: mRNA

A:Residues: 1-638 <RES>

A:Cross-references: GB:M58590; NID:g206721; PIDN:AAA42069.1; PID:g206722

C:Comment: This protein, synthesized in the liver, circulates as a noncovalent complex w

C:Comment: The zymogen is activated by factor XIIa, which cleaves the molecule into a 11k

are linked by one or more disulfide bonds.

C:Genetics:

A:Gene: PK

C:Superfamily: coagulation factor XI; trypsin homology

C:Keywords: blood coagulation; duplication; fibrinolysis; glycoprotein; hydrolase; infla

F:1-19/Domain:signal sequence #status predicted <SIG>
F:20-390/Product: plasma kallikrein heavy chain #status experimental <MAT1>
F:20-103/Domain: apple repeat <AP1>
F:110-199/Domain: apple repeat <AP2>
F:200-289/Domain: apple repeat <AP3>
F:291-380/Domain: apple repeat <AP4>
F:391-638/Product: plasma kallikrein light chain #status experimental <MAT2>
F:391-632/Domain: trypsin homology <TRY>
F:21-104,47-77,51-57,111-194,137-166,141-147,201-284,227-256,231-237,292-375,318-347,
F:127,215,308,453,459,494/Binding site: carboxylate (Asn) (covalent) #status predicted
F:356/Binding site: carboxylate (Asn) (covalent) #status experimental
F:433,453,578/Active site: His, Asp, Ser #status predicted

[illegible]

```

RESULT      3
KOMPFL
plasma kallikrein (EC 3.4.21.34) precursor - mouse
C:Species: Mus musculus (house mouse)
C:Date: 30-Sep-1992 #sequence_revision 30-Sep-1992 #text_change 18-Jun-1999
C:Accession: A36557
R:Seidah, N.G.; Sawyer, N.; Hamelin, J.; Mion, P.; Beaubien, G.; Brachpapa, L.; Rochefort, D.
DNA Cell Biol. 9, 737-748, 1990
A:Title: Mouse plasma kallikrein: cDNA structure, enzyme characterization, and comparison with human plasma kallikrein
A:Reference number: A36557; MUID:91090844
A:Accession: A36557
A:Molecule type: mRNA
A:Residues: 1-638 <SEQ>
A:Cross-references: GB:58588; NID:9200358; PIDD:AAA6393.1; PID:9200359
A:Note: part of this sequence, including the amino ends of both the heavy and light chains, is identical to the sequence of the human plasma kallikrein.
C:Comment: This protein, synthesized in the liver, circulates as a noncovalent complex with the zymogen, which is activated by factor XIIa, which cleaves the molecule into a heavy and a light chain.
C:Linkage: The zymogen is activated by factor XIIa, which cleaves the molecule into a heavy and a light chain.
C:Superfamily: coagulation factor XI; trypsin homology
C:Keywords: blood coagulation; duplication; fibrinolysis; glycoprotein; hydrolase; in
F:1-19/Dom: signal sequence #status predicted <IG>
F:20-39/Product: plasma kallikrein heavy chain #status experimental <HC>
F:20-109/Dom: apple repeat <AP1>
F:110-199/Dom: apple repeat <AP2>
F:200-289/Dom: apple repeat <AP3>
F:291-380/Dom: apple repeat <AP4>
F:391-638/Product: plasma kallikrein light chain #status experimental <LC>
F:391-631/Dom: trypsin homology <TRY>
F:21-104, 47-77, 51-57, 111-194, 137-166, 141-147, 201-284, 227-256, 231-237, 292-375, 318-347,
F:121, 215-308, 396, 494/Binding site: carbohydrate (Asn) (covalent) #status predicted
F:434, 483, 578/Active site: His, Asp, Ser #status predicted

```

Cross-References:

DB 390 RIVGSTNLSLEEMWPQVSLQYKLVISQTHLCGSIITGGRWLTAAHCD - GIPDYDWRIY 448
OY 58 AGILRQSFMFEGYA - GYOVEKYIHPNDSDTKRNDDIALMKLQKLFPENDLKVPCLEPNP 116
Db 449 GGIIISTEITKEPFSSRIKKELIHQEVKVSSEGNIDALIKLPDLANTFERQPICLPSKA 508
OY 117 MMLPEOLCWNISMGATEREKRTSEVLNAKAVLLIEFORCSRVIYNLNLTTPMATICGFL 176
Db 509 DNTTYTCWMTGVGYKRYKEGETQNILLOKATIPPLVPNECCPKY - RDVIYNKOWICAGYK 567

Sequence:

OY 177 QGNVDSCGDGSGGLPVTSKNINMWLIIGDTSNGSGCANARPGVGNYAFETDIYROMRA 236
+ :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|:
Db 568 ECGDACKGDSGCGLPYCKHSGRMQLVTGSWGECGRKDOPGVITYKSEYMDMIILEKTOS 627

RESULT 4

KFHU
coagulation factor Xla (EC 3.4.21.27) precursor [validated] - human
N:Alternate names: antihemophilic factor C; plasma thromboplastin antecedent
C:Species: Homo sapiens (man)
C>Date: 13-Aug-1986 #sequence_revision 26-May-1994 #text_change 08-Dec-2000
A:Accession: A27431; A00920; A37940
R:Sasaki, R.; Davie, E.W.; Chung, D.W.
Biochemistry 26, 7221-7228, 1987
A>Title: Organization of the gene for human factor XI.
A:Reference number: A27431; MUID:88107663
A:Accession: A27431
A:Molecule type: DNA
A:Residues: 1-625 <ASAS>
A:CROSS-references: GB:M18295
A>Note: The sequence shown follows the authors' translation
R:Fujikawa, K.; Chung, D.W.; Hendrickson, L.E.; Davie, E.W.
Biochemistry 25, 2417-2424, 1986
A:Reference number: A00920; MUID:66243360
A:Accession: A00920
A:Molecule type: protein
A:Residues: 1-625 <FDJ>
A:CROSS-references: FJM13142; NID:g182832; PIDN:AAA5487.1; PID:g182833
R:McMullen, B.A.; Fujikawa, K.; Davie, E.W.
Biochemistry 30, 2056-2060, 1991
A>Title: Location of the disulfide bonds in human coagulation factor XI: the presence of
A:Reference number: A37940; MUID:91152017
A:Accession: A37940
A:Molecule type: Protein
A:Residues: 28-33,35-49,'X',51-55,'X',57-63,70-75,'X',77-79,107-109,'X',111-112,132-139,
+ 280-282,'X',284,285-297,313-316,'X',318-319,320-326,'X',328-330,'X',347-349,373,'X',375
C:Comment: The proenzyme consists of two identical chains linked by one or more disulfide
the active site, and a heavy chain, which associates with high molecular weight (HMW) kin
C:Genetics:
A:Gene: GDB:F11
A:CROSS-references: GDB:119891; OMIM:264900
A:Map position: 4q35-q435
A:Introns: 19/1; 73/2; 109/1; 162/2; 199/1; 252/2; 289/1; 343/2; 379/1; 435/2; 494/1; 52
C:Function:
A>Description: catalyzes the proteolytic activation of coagulation factor IX
A:Pathway: blood coagulation intrinsic pathway
C:Superfamily: coagulation factor XI; trypsin homology
C:Keywords: blood coagulation; duplication; glycoprotein; hemophilia C; homodimer; hydr
F:1-18/Domin: signal sequence #status predicted <SIG>
F:19-38//Product: coagulation factor Xla heavy chain #status experimental <HCN>
F:119-108/Domain: apple repeat <AP1>
F:109-198/Domain: apple repeat <AP2>
F:199-288/Domain: apple repeat <AP3>
F:290-379/Domain: apple repeat <AP4>
F:388-635/Product: coagulation factor Xla light chain #status experimental <LCH>
F:388-618/Domain: tryptsin homology <TRY>
F:20-103,514-581,571-599/Disulfide bonds: #status predicted
F:29/Disulfide bonds: Interchain #status experimental
F:946-76,50-56,110-193,136-165,140-146,200-283,226-255,230-236,291-374,317-346,321-327,38
F:90,126,353,450/Binding site: carbohydrate (Asn) (covalent) #status predicted
F:339/Disulfide bonds: Interchain #status predicted

[illegible]

```

Db 341 YPEGIDACQCGSGPFVCEDSISRPRMRLOGIYVMTGCAIAKPGVYTRVSDFREWI 400
Oy 231 YRQMR 235
Db 401 FOAIR 405

RESULT 6
A43090
enteropeptidase (EC 3.4.21.9) precursor - bovine
N:Alternate names: enterokinase
C:Species: Bos primigenius taurus (cattle)
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
C:Accession: A43090; A48874; A61436
R:Klitamoto, Y.; Yuan, X.; Wu, Q.; McCourt, D.W.; Sadler, J.E.
Proc. Natl. Acad. Sci. U.S.A. 91, 7588-7592, 1994
A:Title: Enterokinase, the initiator of intestinal digestion, is a mosaic protease compo
A:Reference number: A43090; MUID:94329561
A:Accession: A43090
A:Status: nucleic acid sequence not shown; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-1035 <RT>
A:Cross-references: GB:U09859; NID:g746410; PIDN:ABA40026.1; PID:g746411
A:Experimental source: small intestine
R:Lavallie, E.R.; Rehemtulla, A.; Racle, L.A.; DiBasilio, E.A.; Ferenz, C.; Grant, K.L.;
J. Biol. Chem. 268, 23311-23317, 1993
A:Title: Cloning and functional expression of a cDNA encoding the catalytic subunit of h
A:Reference number: A48874; MUID:94043122
A:Accession: A48874
A:Molecule type: mRNA
A:Residues: 801-1035 <LAV>
A:Cross-references: GB:L19663; NID:g416131; PIDN:AAA16035.1; PID:g416132
A:Note: parts of this sequence, including the amino end of the mature protein, were conf
R:Light, A.; Janske, H.
J. Protein Chem. 10, 475-480, 1991
A:Title: The amino-terminal sequence of the catalytic subunit of bovine enterokinase.
A:Reference number: A61436; MUID:92189715
A:Accession: A61436
A:Molecule type: protein
A:Residues: 801-807, 'Y', 809-827 <LIG>
C:Comment: The mechanism of association with the membrane of the intestinal brush border
embaine attachment using a signal-anchor sequence.
C:Comment: Conversion from membrane-bound to soluble forms may involve further processin
C:Complex: mature enteropeptidase is variously reported to contain two (heavy and light)
lfiide linked
C:Function:
A:Description: cleaves propeptide from trypsinogen to produce active trypsin
A:Pathway: intestinal digestive hydrolase cascade
C:Superfamily: enteropeptidase; Clr/Cls repeat homology; LDL receptor ligand-binding reg
C:Keywords: glycoprotein; hydrolase; intestine; serine proteinase; transmembrane protein
F:12-38/Domain: transmembrane #status predicted <TMM>
F:52-117/Product: enteropeptidase mini chain #status predicted <KCH>
F:118-800/Product: enteropeptidase heavy chain #status predicted <KCH>
F:199-226/Domain: LDL receptor ligand-binding repeat homology <LDL1>
F:199-226/Domain: LDL receptor ligand-binding repeat homology <LDL1>
F:358-520/Domain: MAM homology <MAM>
F:542-647/Domain: Clr/Cls repeat homology <CLR>
F:659-799/Domain: scavenger receptor cysteine-rich domain homology #status atypical <SRC
F:801-1035/Product: enteropeptidase light chain #status predicted <LCH>
F:801-1030/Domain: trypsin homology <TRY>
F:116,147,170,194,233,263,264,404,456,486,519,550,646,698,722,741,762,864,903,965/Binding sit
F:788-912,826-842,926-993,957-972,983-1011/Disulfide bonds: #status predicted
F:841,892,987/Active site: His, Asp, Ser #status predicted

```

```

Oy 61 LRQSEMFYAGYQVE-----KVISHPNYDSKTKNDNDIAIMKLOKLPTENDLVKPYCLPNP 115
Db 860 HMASN-----TSPQIETRLIDQIVINPHNKRKNNNDIAMHLEMKVNTDYIOPCLPEE 916
Oy 116 GMILOPEOLCWSIGMGATEEKRTSEVLNAKVLLETRQCRSRYVYDNLITPAMICAGF 175
Db 917 NCVFPPGRICSIAGWALYQGSTADVLQEDAVYPLLSNEKCCQCMPEYN-ITENNVCAGY 975
Oy 176 LOGNVDSQCGSGGLVYTSKNIMWLDGTSMSGSCAKAYRPGVGNVLPFDWI 230
Db 976 EAGVDSQCGSGSGPLMCOENNRMLLAGVTSFYOCALPNRGVYARVPFEWI 1030

RESULT 7
A56318
enteropeptidase (EC 3.4.21.9) precursor - human
N:Alternate names: enterokinase
C:Species: Homo sapiens (man)
C:Date: 19-May-1995 #sequence_revision 09-Aug-1996 #text_change 18-Jun-1999
C:Accession: A56318; B43090
R:Klitamoto, Y.; Yuan, X.; Wu, Q.; McCourt, D.W.; Sadler, J.E.
Biochemistry 34, 4562-4568, 1995
A:Title: cDNA sequence and chromosomal localization of human enterokinase, the proteo
A:Reference number: A56318; MUID:95234679
A:Accession: A56318
A:Molecule type: mRNA
A:Residues: 1-1019 <RT>
A:Cross-references: GB:U09860; NID:g746412; PIDN:AAC50138.1; PID:g746413
R:Klitamoto, Y.; Yuan, X.; Wu, Q.; McCourt, D.W.; Sadler, J.E.
Proc. Natl. Acad. Sci. U.S.A. 91, 7588-7592, 1994
A:Title: Enterokinase, the initiator of intestinal digestion, is a mosaic protease co
A:Reference number: A43090; MUID:94329561
A:Accession: B43090
A:Status: nucleic acid sequence not shown
A:Molecule type: mRNA
A:Residues: 749-1019 <KIT>
A:Cross-references: GB:U09860
C:Comment: The mechanism of association with the membrane of the intestinal brush bor
cated below) or with amino-terminal myristoylation of the heavy chain.
C:Genetics:
A:Gene: GDB:PRSS7
A:Cross-references: GDB:384083; OMIM:226200
A:Map position: 21q21-21q21
C:Complex: Mature enteropeptidase is variously reported to contain two (heavy and lig
ed by a disulfide bond. Possibly, conversion from membrane-bound to soluble forms inv
ducts.
C:Function:
A:Description: cleaves activation peptide from trypsinogen to produce active trypsin
A:Pathway: intestinal digestive hydrolase cascade
C:Superfamily: enteropeptidase; Clr/Cls repeat homology; LDL receptor ligand-binding
C:Keywords: glycoprotein; hydrolase; intestine; serine proteinase; transmembrane protein; zymoge
F:1-784/Product: enteropeptidase heavy chain #status predicted <KCH>
F:12-38/Domain: transmembrane #status predicted <TMM>
F:184-221/Domain: LDL receptor ligand-binding repeat homology <LDL1>
F:184-221/Domain: LDL receptor ligand-binding repeat homology <LDL1>
F:342-504/Domain: MAM homology <MAM>
F:526-631/Domain: Clr/Cls repeat homology <CLR>
F:659-799/Domain: scavenger receptor cysteine-rich domain homology #status atypical <
F:785-1019/Product: enteropeptidase light chain #status predicted <LCH>
F:785-1014/Domain: trypsin homology <TRY>
F:116,147,179,328,335,388,440,470,503,534,630,662,706,725,848,887,909,949/Binding sit
F:772-896,810-826,910-977,941-956,967-995/Disulfide bonds: #status predicted
F:825,876,971/Active site: His, Asp, Ser #status predicted

```

```

Query Match 39.7%, Score 523.5; DB 1; Length 1035;
Best Local Similarity 42.1%, Pred. No. 2.1e-39;
Matches 99; Conservative 45; Mismatches 82; Indels 9; Gaps 3;

```

```

Oy 1 RIVGESALPGAMPMQVSLHVNQVHVGSGSITTPMITYTAHACVEKPLNNPMHTAFAGI 60
Db 784 RIVGGSNAKEGAMPVWVGLYGRLLCGASLVSSDWLVSAHCYVGRNLEPSKMTAIIGL 843

```


THIS PAGE BLANK (USPTO)

```

OM protein - protein search, using sw model
Run on:      September 26, 2001, 09:27:26 ; Search time 16.75 Seconds
              (without alignments)
              486.735 Million cell updates/sec

```

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

```
Searched:      93435 seqs, 34255486 residues
Total number of hits satisfying chosen parameters:  93435
```

```
Minimum DB seq length: 0
Maximum DB seq length: 20000000000
```

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : SwissProt_39:*

Pred. NO. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

SUMMARIES

Result	No.	Score	Match	Length	ID	Description
	1	1303	98.8	492	1	TMS2_HUMAN
	2	1105	83.8	490	1	TMS2_MOUSE
	3	645.5	48.9	454	1	TMS3_HUMAN
	4	558.5	42.3	638	1	KAL_HUMAN
	5	543	41.2	638	1	KAL_MOUSE
	6	530	40.2	638	1	KAL_MOUSE
	7	529	40.1	625	1	FALL_HUMAN
	8	528.5	40.1	625	1	HEPS_HUMAN
	9	523.5	39.7	1035	1	ENTK_BOVIN
	10	513.5	38.9	1019	1	ENTK_HUMAN
	11	507.5	38.5	416	1	HEPS_MOUSE
	12	504	38.2	1069	1	ENTK_MOUSE
	13	502.5	38.1	812	1	PLMN_MOUSE
	14	502	38.1	835	1	ST14_HUMAN
	15	499.5	37.9	416	1	HEPS_HUMAN
	16	492	37.3	421	1	ACRO_HUMAN
	17	489	37.1	902	1	ST14_MOUSE
	18	488.5	37.0	1034	1	ENTK_PIG
	19	481.5	36.5	436	1	ACRO_PIG
	20	476.5	36.1	436	1	ACRO_MOUSE
	21	466.5	35.4	431	1	ACRO_RABIT
	22	464	35.2	790	1	PLMN_PIG
	23	460.5	34.9	437	1	ACRO_RAT
	24	458	34.7	810	1	PLMN_MACMU
	25	457.5	34.7	810	1	PLMN_EREDU
	26	454	34.4	343	1	PLMN_SHEEP
	27	451.5	34.2	812	1	PLMN_BOVIN
	28	450.5	34.2	267	1	TRVY_ANOGA
	29	450	34.1	270	1	TRVY_MERUN
	30	449	34.0	275	1	TRV4_ANOGA
	31	449	34.0	343	1	PS68_HUMAN
	32	447	33.9	276	1	MCT6_MOUSE
	33	447	33.9	277	1	TRV2_ANOGA

34	443	33.6	27.4	1	TRV1_ANOGA	P35033	anopheles
35	443	33.6	33.8	1	PLMN_HORSE	P80010	equus caball
36	442.5	33.5	45.8	1	APO4_HUMAN	P08519	homo sapien
37	442	33.5	81.0	1	PLMN_HUMAN	P00747	homo sapien
38	440	33.4	27.4	1	TRIM_RAT	P50343	homo sapien
39	439	33.3	27.3	1	MC17_MOUSE	P00284	mus musculus
40	437	33.1	27.5	1	TRV3_ANOGA	P35033	anopheles g
41	436	33.1	27.3	1	MC17_RAT	P27455	rattus norv
42	436	33.1	33.3	1	PLMN_CANFA	P80009	canis famili
43	435	33.0	24.5	1	TRIM_GANDMO	P80646	gadus morhu
44	429	32.5	25.3	1	CITR_RAT	P07335	rattus norv
45	429	32.5	27.5	1	TRPB_HUMAN	P20231	homo sapien

ALIGNMENTS

	RESULT	1			
TMS2_HUMAN	ID	TMS2_HUMAN	STANDARD:	PRT:	492 AA.
AC	O15313;				
DT	15-JUL-1998	(Rel. 36, Created)			
DT	15-JUL-1998	(Rel. 36, Last sequence update)			
DE	01-OCT-2000	(Rel. 40, Last annotation update)			
GN	TRANSMEMBRANE PROTEASE, SERINE 2 (EC 3.4.21.-).				
OS	TMPSR2.				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.				
OX	NCBT_TaxID=9606;				
RP	[1]				
RX	SEQUENCE FROM N.A.				
RA	MEDLINE=97468144; PubMed=93325052;				
RA	Paoloni-Giacchino A., Chen H., Peltsch M.C., Rossler C., Antonarakis S.E.				
RT	"Cloning of the TMPSR2 gene, which encodes a novel serine protease with transmembrane, LDLRA, and SRCR domains and maps to 21q22.3.";				
RL	Genomics 44:309-320(1997).				
CC	- I - SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN.				
CC	- I - TISSUE SPECIFICITY: EXPRESSED STRONGLY IN SMALL INTESTINE AND WEAKLY IN SEVERAL OTHER TISSUES.				
CC	- I - SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE TRYPSIN FAMILY.				
CC	- I - SIMILARITY: CONTAINS 1 SRCR DOMAIN.				
CC	- I - SIMILARITY: CONTAINS 1 LDL-RECEPTOR CLASS A DOMAIN.				
CC	-----				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see http://www.isb-sib.ch/announce/) or send an email to license@isb-sib.ch .				
CC	-----				
DR	EMBL: U75329; AAC51784.1; ..				
DR	HSSP: P00763; IDPO:				
DR	MIM: 602060;				
DR	InterPro: IPR001254; ..				
DR	InterPro: IPR001314; ..				
DR	InterPro: IPR002172; ..				
DR	pfam: PF00057; 1d1_recept_a; 1.				
DR	pfam: PF00089; trypsin; 1.				
DR	PRINTS: PR00722; CHYMOTRYPSIN.				
DR	PROSITE: PS00134; TRYPsin_HIS; 1.				
DR	PROSITE: PS00135; TRYPsin_SER; 1.				
DR	PROSITE: PS01209; LDLRA_1; 1.				
DR	PROSITE: PS50068; LDLRA_2; 1.				
KW	Hydrolase; Serine protease; Transmembrane; Signal-anchor.				
FT	DOMAIN	1	84	CYTOLASMIC (POTENTIAL).	
FT	TRANSMEM	85	105	SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN) (POTENTIAL).	
FT	EXTRACELLULAR	106	492	EXTRACELLULAR (POTENTIAL).	
FT	DOMAIN	112	150	LDL-RECEPTOR CLASS A.	

```

FT DOMAIN 151 243 SRCR.
FT DOMAIN 256 492 CATALYTIC.
FT ACT_SITE 296 492 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 345 445 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 441 441 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT DISULFID 113 126 BY SIMILARITY.
FT DISULFID 120 139 BY SIMILARITY.
FT DISULFID 133 148 BY SIMILARITY.
FT DISULFID 244 365 BY SIMILARITY.
FT DISULFID 281 297 BY SIMILARITY.
FT DISULFID 410 426 BY SIMILARITY.
FT DISULFID 437 465 BY SIMILARITY.
FT CARBOHYD 213 213 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 249 249 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 492 AA; 53847 MW; 7EAFED18609DDA CRC64;

Query Match 98.8%; Score 1303; DB 1; Length 492;
Best Local Similarity 98.3%; Pred. No. 3.6e-110;
Matches 234; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 RIVGESALPGAMPQVSLHVNVCSSITPEWITVAHCVEKPLNPMHTAFAGI 60
DB 255 RIVGESALPGAMPQVSLHVNVCSSITPEWITVAHCVEKPLNPMHTAFAGI 314
QY 61 LROSFMEYAGAYOEKVISHPNYSKTKNDIALMKLQPLTFNDLVKPVCLPBGMLQ 120
DB 315 LROSFMEYAGAYOEKVISHPNYSKTKNDIALMKLQPLTFNDLVKPVCLPBGMLQ 374
QY 121 PEOLCMTSGMGATEEKGTSVLAANKVLLIETORCNSRYVDNLITPAMICAGFLQGNV 180
DB 375 PEOLCMTSGMGATEEKGTSVLAANKVLLIETORCNSRYVDNLITPAMICAGFLQGNV 434
QY 181 DSCGDSGGPLVTSKNNIMWLGDTSMGSCAKAYRPGVGNVWFTDWIYROMRADG 238
DB 435 DSCGDSGGPLVTSKNNIMWLGDTSMGSCAKAYRPGVGNVWFTDWIYROMKANG 492

RESULT 2
TMS2_MOUSE STANDARD; PRT; 490 AA.
ID TMS2_MOUSE
AC 09JIOB: 09JIC4: 09QYB2:
DT 01-OCT-2000 (Rel. 40, Created)
DT 01-OCT-2000 (Rel. 40, Last sequence update)
DE TRANSMEMBRANE PROTEASE, SERINE 2 (EC 3.4.21.-) (EPITHELIALIN) (PLASMIC
GN TRANSMEMBRANE PROTEIN X).
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BALB/C;
RA Varata M.H., Portari K.S., Kellokumpu S., Kyllonen A.P., Vihko P.T.;
RT Expression of transmembrane serine protease TMPRSS2 in mouse and
RT human tissues.
RL J. Pathol. 191:0-0(2000).
RN [2]
RP SEQUENCE FROM N.A.
RA Han J., Kim S.;
RT Putative transmembrane protease X.
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-BALB/C;
RA Jacquinet E.J., Rao N.V., Rao G.N., Hoidal J.R.;
RT A novel mosaic serine protease, epithelisin.
RL Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.
CC - SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN.
CC - SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC TRYPSIN FAMILY.

```

```

CC - SIMILARITY: CONTAINS 1 SRCR DOMAIN.
CC - SIMILARITY: CONTAINS 1 LDL-RECEPTOR CLASS A DOMAIN.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb.ch/announce/
or send an email to license@isb.ch).
-----
DR EMBL: AF199362; AAF97867.1;
DR EMBL: AF243500; AAF64186.1;
DR EMBL: AF113596; AAF21308.1;
DR HSSP: P00761; IAKS.
DR MCD: MGI:1354381; TmpRSS2.
DR InterPro: IPR001254;
DR InterPro: IPR001314;
DR InterPro: IPR002172;
DR PRINTS: PR00722; CHYMOTRYPSIN.
DR PROSITE: PS00134; TRYPSIN_HIS. 1.
DR PROSITE: PS00135; TRYPSIN_SER. 1.
DR PROSITE: PS01209; LDLRA_1. 1.
DR PROSITE: PS50068; LDLRA_2. 1.
KW Hydroxylase; Serine protease; Transmembrane; Signal-anchor.
FT DOMAIN 1 83 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 84 104 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
FT TRANSMEM 1 83 EXTRACELLULAR (POTENTIAL).
FT DOMAIN 105 490 LDL-RECEPTOR CLASS A.
FT DOMAIN 111 149 SRCR.
FT DOMAIN 150 242 SRCR.
FT DOMAIN 254 490 CATALYTIC.
FT ACT_SITE 294 294 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 343 343 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 439 439 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT DISULFID 112 125 BY SIMILARITY.
FT DISULFID 119 138 BY SIMILARITY.
FT DISULFID 132 147 BY SIMILARITY.
FT DISULFID 243 363 BY SIMILARITY.
FT DISULFID 279 295 BY SIMILARITY.
FT DISULFID 408 424 BY SIMILARITY.
FT DISULFID 435 463 BY SIMILARITY.
FT CARBOHYD 111 111 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 212 212 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 474 474 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CONFLICT 122 122 S -> L (IN REF. 3).
FT CONFLICT 178 178 S -> N (IN REF. 3).
FT CONFLICT 320 320 Y -> H (IN REF. 1).
FT CONFLICT 474 474 N -> D (IN REF. 1).
SQ SEQUENCE 490 AA; 53479 MW; 07DB03E4DBA1A9 CRC64;

Query Match 83.8%; Score 1105; DB 1; Length 490;
Best Local Similarity 84.4%; Pred. No. 2.6e-92;
Matches 200; Conservative 15; Mismatches 22; Indels 0; Gaps 0;

QY 1 RIVGESALPGAMPQVSLHVNVCSSITPEWITVAHCVEKPLNPMHTAFAGI 60
DB 253 RIVGESALPGAMPQVSLHVNVCSSITPEWITVAHCVEKPLNPMHTAFAGI 312
QY 61 LROSFMEYAGAYOEKVISHPNYSKTKNDIALMKLQPLTFNDLVKPVCLPBGMLQ 120
DB 313 LROSFMEYAGAYOEKVISHPNYSKTKNDIALMKLQPLTFNDLVKPVCLPBGMLQ 372
QY 121 PEOLCMTSGMGATEEKGTSVLAANKVLLIETORCNSRYVDNLITPAMICAGFLQGNV 180
DB 373 PEOLCMTSGMGATEEKGTSVLAANKVLLIETORCNSRYVDNLITPAMICAGFLQGNV 432
QY 181 DSCGDSGGPLVTSKNNIMWLGDTSMGSCAKAYRPGVGNVWFTDWIYROMRAD 237
DB 433 DSCGDSGGPLVTSKNNIMWLGDTSMGSCAKAYRPGVGNVWFTDWIYROMKANG 489

```



```

RESULT 3
TMS3_HUMAN STANDARD: PRT; 454 AA.
AC P57727.
DT 01-OCT-2000 (rel. 40, Created)
DT 01-OCT-2000 (rel. 40, Last sequence update)
DT 01-OCT-2000 (rel. 40, Last annotation update)
DE TRANSMEMBRANE PROTEASE, SERINE 3 (EC 3.4.21.-) (SERINE PROTEASE
TAD-12) (TUMOR ASSOCIATED DIFFERENTIALLY-EXPRESSED GENE-12 PROTEIN).
GN TMPRSS3 OR TADG12 OR ECHOS1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORMS A AND TRUNCATED).
RC TISSUE-ovarian carcinoma;
RA MEDLINE=20521358; PubMed=11068177;
RA Underwood L.J., Shigemasa K., Tanimoto H., Beard J.B., Schneider E.N.,
RA Wang Y., Parmley T.H., O'Brien T.J.;
RA "Ovarian tumor cells express a novel multi-domain cell surface serine
RT protease."
RL Biochim. Biophys. Acta 1502:337-350(2000).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORMS A; B/C AND D).
RX PubMed=11137999;
RA Scott H.S., Kudoh J., Mattenhofer M., Shibuya K., Berry A., Chrast R.,
RA Gulponi M., Wang J., Kawasaki K., Asakawa S., Minoshima S.,
RA Yonuss F., Mehl S.O., Radhakrishna U., Pappasavas M.P., Gehrig C.,
RA Rossler C., Korostilovsky M., Gal A., Shimizu N., Bonne-Tamir B.,
RA Antonarakis S.E.;
RA "Insertion of beta-satellite repeats identifies a transmembrane
RT protease causing both congenital and childhood onset autosomal
RT recessive deafness."
RL Nat. Genet. 27:59-63(2001).
CC -1 SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN (POTENTIAL).
CC -1 ALTERNATIVE PRODUCTS: 4 ISOFORMS; A (SHOWN HERE), B/C, D AND
CC TRUNCATED/TAD-12; ARE PRODUCED BY ALTERNATIVE SPLICING.
CC -1 TISSUE SPECIFICITY: EXPRESSED IN MANY TISSUES. THE TRUNCATED
CC ISOFORM IS FOUND AT INCREASED LEVELS IN SOME CARCINOMAS.
CC -1 DISEASE: DEFECTS IN TMPRSS3 ARE A CAUSE OF TWO FORMS OF AUTOSOMAL
CC NEUROSENSORY CHILDHOOD-ONSET FORMS OF DEAFNESS, DFNB8 AND DFNB10.
CC -1 SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC TRYPSIN FAMILY.
CC -1 SIMILARITY: CONTAINS 1 SRCR DOMAIN.
CC -1 SIMILARITY: CONTAINS 1 LDL-RECEPTOR CLASS A DOMAIN.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AF201380; AAG37012.1; -
CC EMBL: AB038157; BAB20077.1; -
CC EMBL: AB038158; BAB20078.1; -
CC EMBL: AB038159; BAB20079.1; -
CC EMBL: AB038160; BAB20080.1; -
CC MIM: 605511; -
CC MIM: 601072; -
CC MIM: 605316; -
CC PROSITE: PS01209; LDLRA_1; 1.
CC PROSITE: PS50068; LDLRA_2; 1.
CC PROSITE: PS50287; SRCR_2; 1.
CC PROSITE: PS00134; TRYPSIN_HIS; 1.
CC PROSITE: PS00135; TRYPSIN_SER; 1.
CC K1 Hydrolyase: Serine protease; Transmembrane; Signal-anchor; Deafness;
CC Alternative splicing
CC DOMAIN 1 28 CYTOPLASMIC (POTENTIAL).
CC TRANSMEM 29 49 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
CC (POTENTIAL).

```

```

FT DOMAIN 50 454 EXTRACELLULAR (POTENTIAL).
FT DOMAIN 72 108 LDL-RECEPTOR CLASS A.
FT DOMAIN 109 205 SRCR.
FT DOMAIN 217 454 CATALYTIC.
FT ACT_SITE 257 257 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 304 304 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 401 401 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT DISULFID 73 85 BY SIMILARITY.
FT DISULFID 79 98 BY SIMILARITY.
FT DISULFID 92 107 BY SIMILARITY.
FT DISULFID 207 324 BY SIMILARITY.
FT DISULFID 242 258 BY SIMILARITY.
FT DISULFID 397 425 BY SIMILARITY.
FT CARBOHYD 221 221 N-LINKED (GLCNAC. .) (POTENTIAL).
FT VARSPPLIC 1 127 MISSING (IN ISOFORM B/C).
FT VARSPPLIC 318 454 MISSING (IN ISOFORM A).
FT FT
FT FT
FT FT
FT VARSPPLIC 261 293
FT FT
FT VARSPPLIC 294 454
FT CONFLICT 46 54 MISSING (IN TRUNCATED ISOFORM).
FT CONFLICT 90 90 LKFFPIVI -> FEVSGSSSL (IN REF. 1).
FT CONFLICT 350 350 A -> T (IN REF. 1).
FT CONFLICT 369 395 MISSING (IN REF. 1).
FT CONFLICT 427 427 ICHNRDVGITISPSMLCAGYLTGYD -> DLQPOGRVWR
FT CONFLICT 427 427 HHLPLALRGLPDGMWN (IN REF. 1).
FT SEQUENCE 454 AA; 49404 MW; 57ECC3678FD6AF CRC64;

Query Match 48.9%; Score 645.5; DB 1; Length 454;
Best Local Similarity 52.7%; Pred. No. 6,3e-51;
Matches 126; Conservative 30; Mismatches 78; Indels 5; Gaps 4;

QY 1 RIYGGESALPGAMPVOYSLHVNWHVCGGSIIPENIVTAHGVENPLNPMHTAFAGI 60
DB 216 RIYGGNSLSLQMPWQSLQFQGIHLGGSVITPLMTTAAHCV-YDLYLPKSMITQVGL 274
QY 61 LRQSFMEYGA-GYQVEKVISHPYDSTRKNNIDALMLKQLETFNDLVKVCPLPFGMML 119
DB 275 V-SLIDNPAPSHLVKVEIYVHSKYKPRGLNDIAKMLAGPLFENMIQVCLPSEENE 332
QY 120 QPQLCHISGKATEE-KGRTSEVLAQKVLLETQRCNSRYVDNLITPAMICAGFLQ 178
DB 333 PDKVCWTSQGWATEDEGAGDAPVLNHAAPVLISNKCINHRDVGIGIISPSMLCAGYLTG 392
QY 179 NVDSQGDSCGPIVTSKNNIMWLTIGDTSWMSGCAKARVPCVGNVWFTMYIRORAD 237
DB 393 GVDSQGDSCGPIVQGERLMLKVGATSFQICAEVKNKPGVYTRVTSFLDMIHQMERD 451

RESULT 4
KAL_HUMAN STANDARD: PRT; 638 AA.
AC P03952.
DT 23-OCT-1986 (rel. 02, Created)
DT 23-OCT-1986 (rel. 02, Last sequence update)
DT 01-OCT-2000 (rel. 40, Last annotation update)
DE PLASMA KALLIKREIN PRECURSOR (EC 3.4.21.34) (PLASMA PREKALLIKREIN)
DE (KININOGENIN) (FLETCHER FACTOR).
GN KLK3.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=86243359; PubMed=3521732;
RA Chung D.W., Fujikawa K., McMullen B.A., Davie E.W.;
RA "Human plasma prekallikrein, a zymogen to a serine protease that

```


CC HMW KININOGEN AND MAY ALSO PLAY A ROLE IN THE RENIN-ANGIOTENSIN
 CC SYSTEM BY CONVERTING PRORENIN INTO RENIN.
 CC -1- SUBUNIT: THE ZYMOGEN IS ACTIVATED BY FACTOR XIIA, WHICH CLEAVES
 CC THE MOLECULE INTO A LIGHT CHAIN, WHICH CONTAINS THE ACTIVE SITE,
 CC AND A HEAVY CHAIN, WHICH ASSOCIATES WITH HMW KININOGEN. THESE
 CC CHAINS ARE LINKED BY ONE OR MORE DISULFIDE BONDS.
 CC -1- SIMILARITY: CONTAINS 4 APPLE DOMAINS.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
 CC TRYPSIN FAMILY. PLASMA KALLIKREIN SUBFAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: M62357; AAA74563.1; -
 CC DR EMBL: M62358; AAA74563.1; JOINED.
 CC DR EMBL: M62346; AAA74563.1; JOINED.
 CC DR EMBL: M62347; AAA74563.1; JOINED.
 CC DR EMBL: M62349; AAA74563.1; JOINED.
 CC DR EMBL: M62350; AAA74563.1; JOINED.
 CC DR EMBL: M62351; AAA74563.1; JOINED.
 CC DR EMBL: M62352; AAA74563.1; JOINED.
 CC DR EMBL: M62353; AAA74563.1; JOINED.
 CC DR EMBL: M62354; AAA74563.1; JOINED.
 CC DR EMBL: M62355; AAA74563.1; JOINED.
 CC DR EMBL: M62356; AAA74563.1; JOINED.
 CC DR EMBL: M30282; AAA74563.1; -
 CC DR EMBL: M58590; AAA74563.1; -
 CC DR PIR: A39180; KORTPL.
 CC HSSP: P00750; IRTF.
 CC MEROPS: S01.212; -
 CC DR InterPro: IPR000177; -
 CC DR InterPro: IPR001254; -
 CC DR InterPro: IPR001314; -
 CC DR InterPro: IPR003014; -
 CC DR Pfam: PF00024; PAN; 4.
 CC Pfam: PF00089; Trypsin; 1.
 CC PRINTS: PRO0005; APPLIEDOMAIN.
 CC PRINTS: PRO0722; CHYMOTRYPSIN.
 CC DR PROSITE: PS00134; TRYPSIN_HIS; 1.
 CC DR PROSITE: PS00135; TRYPSIN_SER; 1.
 CC DR PROSITE: PS00495; APPLE; 4.
 CC DR Hydroxylase; Serine protease; Glycoprotein; Plasma; Zymogen; Signal;
 CC Fibrinolysis; Blood coagulation; Inflammatory response; Liver;
 CC Repeat.
 CC KW
 CC FT SIGNAL 1 19
 CC FT CHAIN 20 390
 CC FT CHAIN 391 638
 CC FT DOMAIN 20 105
 CC FT DOMAIN 110 195
 CC FT DOMAIN 200 285
 CC FT DOMAIN 291 376
 CC FT DOMAIN 389 621
 CC FT CARBOHYD 127 127
 CC FT CARBOHYD 308 308
 CC FT CARBOHYD 396 396
 CC FT CARBOHYD 453 453
 CC FT CARBOHYD 494 494
 CC FT ACT_SITE 434 434
 CC FT ACT_SITE 483 483
 CC FT ACT_SITE 578 578
 CC FT ACT_SITE 77 77
 CC FT DISULFID 51 51
 CC FT DISULFID 111 111
 CC FT DISULFID 137 137
 CC FT DISULFID 141 141
 CC FT DISULFID 201 284
 CC FT DISULFID 227 256

FT DISULFID 231 237 BY SIMILARITY.
 FT DISULFID 292 375 BY SIMILARITY.
 FT DISULFID 316 347 BY SIMILARITY.
 FT DISULFID 322 347 BY SIMILARITY.
 FT DISULFID 340 345 BY SIMILARITY.
 FT DISULFID 383 503 BY SIMILARITY.
 FT DISULFID 419 435 BY SIMILARITY.
 FT DISULFID 517 584 BY SIMILARITY.
 FT DISULFID 548 563 BY SIMILARITY.
 FT DISULFID 574 602 BY SIMILARITY.
 SQ SEQUENCE 638 AA; 71273 MW; 454BB27E8CA8F88 CRC64;
 Query Match 41.2%; Score 543; DB 1; Length 638;
 Best Local Similarity 42.9%; Pred. No. 1,6e-41;
 Matches 103; Conservative 46; Mismatches 85; Indels 6; Gaps 4;
 QY 1 RIVGESALPGAMPQVSLHVN---HVGSGSILPPEVITAAHCEKPLNNPMTAF 57
 DB 390 RIVGNTSSIGEMPMQVSLQVLSNHHGSGSILGRWILTAHCFD-GIPYDWRIT 448
 QY 58 AGILROSEMFYAGY-OVERKIVSHPNYDSKTKNNIDALMKLQPLTFNDLVKPCLPNG 116
 DB 449 GGILNLSKITNKTPFSIKELIHOKYKMGSGSYDIALIKLQPLNVTREFOKICLPSRA 508
 QY 117 MNLQPEQLWISGGAETEEKSTSEVLNAKVLLEFORNSVYVNDLITPAMICAGFL 176
 DB 509 DNTITTCWVTGMYGTKEGTONILQKATPLVPEEOKRY-RDYVTKOMICAGK 567
 QY 177 QGNVSCQDGGSGPLVTSKNNIMLIGDTSWGSCKAKAPRYGVGNWFTDVIYQMR 236
 DB 568 EGGIDACKGDSGSPVCKHSGRWQVIGITSMGSCARKRQPGYTVAVETIDILEKIOS 627
 RESULT 6
 KAL_MOUSE STANDARD; PRT; 638 AA.
 AC P26262;
 DT 01-MAY-1992 (Rel. 22, Created)
 DT 01-MAY-1992 (Rel. 22, Last sequence update)
 DT 01-OCT-2000 (Rel. 40, Last annotation update)
 DE PLASMA KALLIKREIN PRECURSOR (EC 3.4.21.34) (PLASMA PREKALLIKREIN)
 DE (KININOGENIN) (FLETCHER FACTOR).
 GN KLK3 OR PK.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 ON NCBI_Taxid-10090;
 RN [1]
 RP SEQUENCE FROM N.A. AND PARTIAL SEQUENCE.
 RC STRAIN-BALB/C; TISSUE-Liver.
 RX MEDLINE-91090844; PubMed-2264928;
 RA Seidah N.G., Sawyer N., Hamelin J., Mion P., Beaubien G.,
 RA Brachpapa L., Rochemont J., Mbikay M., Chretien M.,
 RT "Mouse plasma kallikrein: cDNA structure, enzyme characterization,
 RT and comparison of protein and mRNA levels among species".
 RL DNA Cell Biol. 9:737-748(1990).
 CC -1- FUNCTION: THE ENZYME CLEAVES LYS-ARG AND ARG-SER BONDS. IT
 CC ACTIVATES, IN A RECIPROCAL REACTION, FACTOR XII AFTER ITS BINDING
 CC TO A NEGATIVELY CHARGED SURFACE. IT ALSO RELEASES BRAKININ FROM
 CC HMW KININOGEN AND MAY ALSO PLAY A ROLE IN THE RENIN-ANGIOTENSIN
 CC SYSTEM BY CONVERTING PRORENIN INTO RENIN.
 CC -1- SUBUNIT: THE ZYMOGEN IS ACTIVATED BY FACTOR XIIA, WHICH CLEAVES
 CC THE MOLECULE INTO A LIGHT CHAIN, WHICH CONTAINS THE ACTIVE SITE,
 CC AND A HEAVY CHAIN, WHICH ASSOCIATES WITH HMW KININOGEN. THESE
 CC CHAINS ARE LINKED BY ONE OR MORE DISULFIDE BONDS.
 CC -1- SIMILARITY: CONTAINS 4 APPLE DOMAINS.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
 CC TRYPSIN FAMILY. PLASMA KALLIKREIN SUBFAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its

FT	DISULEPID	416	432	CHAINS.
FT	DISULEPID	514	581	
FT	DISULEPID	545	560	
FT	DISULEPID	571	599	
FT	VARIANT	301	301	F -> L (IN FIL DEFICIENCY).
FT	CONFLICT	226	226	/FRTD-VAR.006622.
FT	SEQUENCE	625 AA;	70109 MW;	C->S (IN REF.2).
				147AF94B7709EB8 CRC64;
	Query Match	40.1%;	Score 529;	DB 1; Length 625;
	Best Local Similarity	42.2%;	Pred. No.2,9e-40;	
	Matches 103;	Conservative 43;	Mismatches 84;	Indels 14; Gaps 5;
QY	1 RIVGESALPGAMFQVSLHAVON--- <td>57</td> <td></td> <td></td>	57		
Db	387 RIVGTSVREEMQVQLHTTSPQRLHCGSITGNOMITLAHCF-YGVESPILRVY	445		
QY	58 AGILRSPM-----FYGAGYOVERKVISHPNDTSKTKNNDIAMKOKLTENDLKRPVCL	112		
Db	446 SGLINOSIEIKEDTSEFFG---VQELTIIDQYKMAESGYDILKLETTVNYVDSQRPICL	501		
QY	113 PNPGMALPEOLCWINSGWATEEKGKTSSEVLNAAKVLLEIETORCNSRYVDNLITPAMIC	172		
Db	502 PSKGDNRVNYTDCWTSQWGYRKLNRDKIGNTLQKAKIPLVTNBECCQRY-RGHKITHKMIC	560		
QY	173 AGFLQGVNDSCQGSQSGPLVTSKNNIMLIGDTISGSCAKAIPRGVYGNVAVFTDNIYR	232		
Db	561 AGYREGGADCKGDSGGLSCKHNEVWMLVGITSWGECQAOREPGVYTVNVEYVDILLE	620		
QY	233 QMRA 236			
Db	621 KTOA 624			
	RESULT. 8			
	HEPS_HUMAN			
ID	HEPS_HUMAN	STANDARD;	PRT;	417 AA.
AC	P05981;			
DT	01-NOV-1988 (Rel. 09, Created)			
DT	01-NOV-1988 (Rel. 09, Last sequence update)			
DT	01-NOV-1997 (Rel. 35, Last annotation update)			
DE	SRINE PROTEASE HEPSIN (EC 3.4.21.-) (TRANSMEMBRANE PROTEASE, SERINE			
GN	1).			
OS	HPN OR TMPRSS1.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.			
OX	NCBI_TaxId=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=Liver.			
RX	MEDLINE=88209431; PubMed=2835076;			
RA	Leytus S.P., Loeb K.R., Hagen F.S., Kurachi K., Davie E.W.;			
RT	"A novel trypsin-like serine protease (hepsin) with a putative			
RT	transmembrane domain expressed by human liver and hepatoma cells."			
RL	Biochemistry 27:1067-1074(1988).			
RN	[2]			
RP	CHARACTERIZATION.			
RX	MEDLINE=91358502; PubMed=1885621;			
RA	Tsuji A., Torres-Rosado A., Arai T., le Beau M.M., Lemons R.S.,			
RT	Chou S.H., Kurachi K.;			
RT	"Hepsin, a cell membrane-associated protease. Characterization,			
RT	tissue distribution, and gene localization.";			
RL	J. Biol. Chem. 266:16948-16953(1991).			
RN	[3]			
RP	CHARACTERIZATION.			
RX	MEDLINE=93348237; PubMed=8346233;			
RA	Torres-Rosado A., O'Shea K.S., Tsuji A., Chou S.H., Kurachi K.;			
RT	"Hepsin, a putative cell-surface serine protease, is required for			
RT	mammalian cell growth.";			
RL	Proc. Natl. Acad. Sci. U.S.A. 90:7181-7187(1993).			

```

CC -1- FUNCTION: PLAYS AN ESSENTIAL ROLE IN CELL GROWTH AND MAINTENANCE
CC OF CELL MORPHOLOGY.
CC -1- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN.
CC -1- TISSUE SPECIFICITY: PRESENT IN MOST TISSUES, WITH THE HIGHEST
CC LEVEL IN LIVER.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC TRYPsin FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: M18930; AAA6013.1; -.
CC EMBL: X07732; CA30558.1; -.
CC EMBL: X07002; CA30058.1; -.
CC PIR: S00845; S00845.
CC HSSP: P00763; IDPO.
CC MEROPS: S01.224; -.
CC MIM: 142440; -.
CC InterPro: IPR001354; -.
CC InterPro: IPR001314; -.
CC Pfam: PF00089; Trypsin; 1.
CC PRINTS: PR00722; CHYMOTRYPSIN.
CC PROSITE: PS00134; TRYPsin_HIS; 1.
CC PROSITE: PS00135; TRYPsin_SER; 1.
CC K1 Hydrolyase; Serine protease; Transmembrane; Signal-anchor.
CC CHAIN 1 162 NON-CATALYTIC CHAIN (POTENTIAL).
CC CHAIN 163 417 CATALYTIC CHAIN (POTENTIAL).
CC DOMAIN 1 17 CYTOPLASMIC (POTENTIAL).
CC TRANSMEM 18 44 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
CC (POTENTIAL).
CC -----
CC DOMAIN 45 417 EXTRACELLULAR (POTENTIAL).
CC ACT_SITE 163 417 CATALYTIC.
CC ACT_SITE 203 203 CHARGE RELAY SYSTEM (BY SIMILARITY).
CC ACT_SITE 257 257 CHARGE RELAY SYSTEM (BY SIMILARITY).
CC ACT_SITE 353 353 CHARGE RELAY SYSTEM (BY SIMILARITY).
CC DISULFID 153 277 INTERCHAIN (BY SIMILARITY).
CC DISULFID 188 204 BY SIMILARITY.
CC DISULFID 322 338 BY SIMILARITY.
CC DISULFID 349 381 BY SIMILARITY.
CC CARBOHYD 112 112 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC SEQUENCE 417 AA; 45011 MW; B2086FF61E551D7 CRC64.

```

```

Query Match 40.1%; Score 528.5; DB 1; Length 417;
Best Local Similarity 41.2%; Pred. No. 2e-40; Indels 11; Gaps 3;
Matches 101; Conservative 41; Mismatches 92;

```

```

OY 1 RIVGESALPGAMPQVSLAHVONVHVGSGSITPEWITVAACVEKPLNNPMHTAFAGI 60
    ||||| 1 ||||| 1 ||||| 1 ||||| 1 ||||| 1 ||||| 1 |||||
DB 169 RIVGSRDRLSLGMRPNQVSLRDCAHLCCGSLSGDWVLTALACFPERRNVLRSRWRFAGA 221
    ||||| 1 ||||| 1 ||||| 1 ||||| 1 ||||| 1 ||||| 1 |||||
OY 61 LROSTMETGAGYQVERKVISHPNY-----DSKTKNDALMLKQPLFNLDLVKPCVLP 114
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 222 VAQASR-HGLQGVQAVVYHGGLYLPFRDPNSSENSNDIALVHLSPLPLEYIQPCIPA 280
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 115 PCMMIOPRDLCSMGWGTAEKGTSEVYLNAAKVLLETRQNSRYVNLTPMATICG 174
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 281 AGQALVLDGICITVGTWGMNTQYGGQAGVLEQEARVPLISNDVCGADFGNQIRPKMFCAG 340
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 175 FLQGVNDSQCGSGGSPV---TSKNINIMLIGDTSWGSAGCAKAYRPGVYGVNWFDTMI 230
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 341 YREGSIDACGSGSPFCVCEDSISRTPRRLCGIYVSWGTCALAKPQGYTVYSDFREWI 400
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 231 YRQMR 235
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 401 FOAIR 405

```

```

RESULT 9
ENTR_BOVIN STANDARD: PRT; 1035 AA.
AC P98072;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE ENTEROPEPTIDASE PRECURSOR (EC 3.4.21.9) (ENTEROKINASE).
GN PRS7 OR ENTK.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidea;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
ON [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC TISSUE-Duodenum;
RX MEDLINE-94329561; PubMed-8052624;
RA Kitamoto Y., Yuan X., Wu Q., McCourt D.W., Sadler J.E.;
RT "Enterokinase, the initiator of intestinal digestion, is a mosaic
RT protease composed of a distinctive assortment of domains.";
RL Proc. Natl. Acad. Sci. U.S.A. 91:7588-7592(1994).
RN [2]
RP SEQUENCE OF 801-1035 FROM N.A., AND PARTIAL SEQUENCE.
RX MEDLINE-94043122; PubMed-8226855;
RA Lavallie E.R., Rehmertulla A., Racie L.A., Dblasio E.A.,
RA Ferenz C., Grant K.L., Light A., McCoy J.M.;
RT "Cloning and functional expression of a cDNA encoding the catalytic
RT subunit of bovine enterokinase.";
RL J. Biol. Chem. 268:23311-23317(1993).
RN [3]
RP SEQUENCE OF 801-827.
RC TISSUE-Intestine;
RX MEDLINE-92189715; PubMed-1799406;
RA Light A., Janska H.;
RT "The amino-terminal sequence of the catalytic subunit of bovine
RT enterokinase.";
RL J. Protein Chem. 10:475-480(1991).
CC -1- FUNCTION: RESPONSIBLE FOR INITIATING ACTIVATION OF PANCREATIC
CC PROTEOLYTIC PROENZYMES (TRYPsin, CHYMOTRYPSIN AND CARBOXYPEPTIDASE
CC A). IT CATALYZES THE CONVERSION OF TRYPsinIN TO TRYPsin WHICH IN
CC TURN ACTIVATES OTHER PROENZYMES INCLUDING CHYMOTRYPSINOGEN,
CC PROCARBOXYPEPTIDASES, AND PROELASTASES.
CC -1- CATALYTIC ACTIVITY: SELECTIVE CLEAVAGE OF 6-LYS-1-ILE-7 BOND IN
CC TRYPsinOGEN.
CC -1- SUBUNIT: HETERODIMER OF A CATALYTIC (LIGHT) CHAIN AND A
CC MULTIDOMAIN (HEAVY) CHAIN LINKED BY A DISULFIDE BOND.
CC -1- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN (PROBABLE).
CC -1- TISSUE SPECIFICITY: INTESTINAL BRUSH BORDER.
CC -1- PTM: THE CHAINS ARE DERIVED FROM A SINGLE PRECURSOR THAT IS
CC CLEAVED BY A TRYPsin-LIKE PROTEASE.
CC -1- SIMILARITY: CONTAINS 2 LDL-RECEPTOR CLASS A DOMAINS.
CC -1- SIMILARITY: CONTAINS 2 CUB DOMAINS.
CC -1- SIMILARITY: CONTAINS 1 SRCR DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 MAM DOMAIN.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC TRYPsin FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: U09659; AA40026.1; -.
CC EMBL: L19663; AA16035.1; -.
CC PIR: A61436; A61436.
CC HSSP: P00763; IDPO.
CC MEROPS: S01.156; -.
CC InterPro: IPR000082; -.
CC InterPro: IPR000859; -.

```

DR InterPro: IPR000998; -
 DR InterPro: IPR001190; -
 DR InterPro: IPR001254; -
 DR InterPro: IPR001314; -
 DR InterPro: IPR002172; -
 DR Pfam: PF00431; CUB; 2.
 DR Pfam: PF00629; MAM; 1.
 DR Pfam: PF01390; SEA; 1.
 DR Pfam: PF00530; SRCR; 1.
 DR Pfam: PF00057; Ldl_recept_a; 2.
 DR Pfam: PF00089; trypsin; 1.
 DR PRINTS: PRO0722; CHYMOTRYPSIN.
 DR PROSITE: PS00134; TRYPSIN_HIS; 1.
 DR PROSITE: PS00135; TRYPSIN_SER; 1.
 DR PROSITE: PS01180; CUB; 2.
 DR PROSITE: PS00740; MAM; 1; 1.
 DR PROSITE: PS50060; MAM; 2; 1.
 DR PROSITE: PS01209; LDLRA_1; 2.
 DR PROSITE: PS50068; LDLRA_2; 2.
 KW Signal-anchor: Glycoprotein; Myristate; Hydrolase;
 KM Serine protease; Zymogen; Transmembrane; Repeat; Alternative splicing.
 FT CHAIN 1 800
 FT CHAIN 801 1035
 FT TRANSMEM 19 47
 FT DOMAIN 197 238
 FT DOMAIN 240 350
 FT DOMAIN 358 520
 FT DOMAIN 540 650
 FT DOMAIN 657 787
 FT DOMAIN 694 841
 FT ACT_SITE 841 892
 FT ACT_SITE 892 987
 FT ACT_SITE 987 987
 FT LIPID 2
 FT DISULFID 199 212
 FT DISULFID 206 225
 FT DISULFID 219 236
 FT DISULFID 659 671
 FT DISULFID 666 684
 FT DISULFID 678 693
 FT DISULFID 788 826
 FT DISULFID 826 842
 FT DISULFID 926 933
 FT DISULFID 957 972
 FT DISULFID 983 1011
 FT CARBOHYD 116 116
 FT CARBOHYD 147 147
 FT CARBOHYD 170 170
 FT CARBOHYD 194 194
 FT CARBOHYD 233 233
 FT CARBOHYD 263 263
 FT CARBOHYD 264 264
 FT CARBOHYD 404 404
 FT CARBOHYD 456 456
 FT CARBOHYD 486 486
 FT CARBOHYD 519 519
 FT CARBOHYD 550 550
 FT CARBOHYD 646 646
 FT CARBOHYD 698 698
 FT CARBOHYD 722 722
 FT CARBOHYD 741 741
 FT CARBOHYD 762 762
 FT CARBOHYD 864 864
 FT CARBOHYD 903 903
 FT CARBOHYD 965 965
 FT VARSPLIC 166 192
 FT CONFLICT 808 808
 SQ SEQUENCE 1035 AA; 114887 MW; E207970B08296E13 CNC64;

Query Match 39.7%; Score 523.5; DB 1; Length 1035;
 Best Local Similarity 42.1%; Pred. No. 1,6e-39;
 Matches 99; Conservative 45; Mismatches 82; Indels 9; Gaps 3;

OY 1 RIVGESALPGAMPQVSLHVNHWCGSSITPEWIVTAHGVCKEPLINPMHMTAFAGI 60
 DB 800 KIVGSDSRREGAMPWYVALYEDDOVCAGSLVSRDVLVAHCVGRNMPSPKKAVALGL 859
 OY 61 LROSFMPYAGAYOVE-----KVISHPNYDSKTKNNIDALMKLOKPLFNDLVPCVLPNP 115
 DB 860 HMASNL---TSPQLETRLDIQIVINPHYKRRKNNIDAMHMLEKKNYTDYIDPCLPEE 916
 OY 116 GMLLOPEQLCMISGWGATEEKGTSYVLNAKYLLEIETORCNSRYVYDNLITPAMICAGF 175
 DB 917 NOVEPPGRICISAGWALLIYGSGTADVLQADADYPLLSNEKCOQOMPEYN-ITENMYCAGY 975
 OY 176 LOGNVDSOCGDSGGLPYTSKNNITMWLIGDTSKSGCAKARPRVGVNVMFTWMI 230
 DB 976 EAGGVDSOQDSGSGPLMCQENNRMLLAGVYSFGYQCALPNRPGVYARVPRTEMI 1030
 RESULT 10
 ID ENTK_HUMAN STANDARD; PRT; 1019 AA.
 AC P98073;
 DT 01-FEB-1996 (Rel. 33, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 01-OCT-2000 (Rel. 40, Last annotation update)
 DE ENTEROPEPTIDASE PRECURSOR (EC 3.4.21.9) (ENTEROKINASE).
 GN PRS7 OR ENTK.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiinae; Homo.
 OX NCBI_Taxid=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Duodenum;
 RX MEDLINE=95234679; PubMed=7718557;
 RA Kitamoto Y., Veille R.A., Donis-Keller H., Sadler J.E.;
 RT "cDNA sequence and chromosomal localization of human enterokinase,
 the proteolytic activator of trypsinogen.";
 RL Biochemistry 34:4562-4568(1995).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Holzinger A., Buck C., Maier E.M., Mayerhofer P.U., Roscher A.A.,
 RA Sadler J.E., Hadorn H.B.;
 RT "Genomic organization of the human enteropeptidase.";
 RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=20289799; PubMed=10830953;
 RA Hattori M., Fujiyama A., Taylor T.D., Watanabe H., Yada T.,
 RA Park H.-S., Toyoda A., Ishii K., Notoki Y., Choi D.-K., Soeda E.,
 RA Ohki M., Takagi T., Sakaki Y., Taudien S., Blechschmidt K., Polley A.,
 RA Menzel U., Delabar J., Kumpf K., Lehmann R., Patterson D.,
 RA Reichwald K., Rump A., Schillhabel M., Schudy A., Zimmermann W.,
 RA Reichenbach K., Kudoh J., Shibuya K., Kawasaki K., Asakawa S.,
 RA Shintani A., Sasaki T., Nagamine K., Mitsuyama S., Antonarakis S.E.,
 RA Minoshima S., Shimizu N., Nordstok G., Hornischer K., Brandt P.,
 RA Scharfe M., Schoen O., Desario A., Reichelt J., Kauer G., Bloecker H.,
 RA Ramser J., Beck A., Klages S., Hennig S., Rieselmann L., Dagand E.,
 RA Wehrmeyer S., Borzym K., Gardiner K., Nizetic D., Francis F.,
 RA Lehrach H., Reinhardt R., Yaspo M.-L.;
 RT "The DNA sequence of human chromosome 21.";
 RL Nature 405:311-319(2000).
 RN [4]
 RP SEQUENCE OF 749-1019 FROM N.A.
 RC TISSUE=Duodenum;
 RX MEDLINE=94329561; PubMed=8052624;
 RA Kitamoto Y., Yuan X., Wu O., McCourt D.W., Sadler J.E.;
 RT "Enterokinase, the initiator of intestinal digestion, is a mosaic
 protease composed of a distinctive assortment of domains.";
 RL Proc. Natl. Acad. Sci. U.S.A. 91:7588-7592(1994).
 CC FUNCTION: RESPONSIBLE FOR INITIATING ACTIVATION OF PANCREATIC
 CC PROTEOLYTIC PRENEZYMS (TRYPSIN, CHYMOTRYPSIN AND CARBOXYPEPTIDASE
 CC A). IT CATALYZES THE CONVERSION OF TRYPSINOGEN TO TRYPSIN WHICH IN


```

RESULT 11
HEPS_MOUSE STANDARD: PRT; 416 AA.
AC 035453;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE SERINE PROTEASE HEPsin (EC 3.4.21.-).
GN HPSN.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver; PubMed=9395459;
RA MEDLINE=98058912; Haakma C., Tomasek J.J., Howard E.W.;
RA Vu T.-K.H., Liu R.W., Haakma C., Tomasek J.J., Howard E.W.;
RT Identification and cloning of the membrane-associated serine
RT protease, hepsin, from mouse preimplantation embryos.
RT J. Biol. Chem. 272:31315-31320(1997).
RL J. Biol. Chem. 272:31315-31320(1997).
CC - FUNCTION: PLAYS AN ESSENTIAL ROLE IN CELL GROWTH AND MAINTENANCE
CC OF CELL MORPHOLOGY.
CC - SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN.
CC - SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC TRYPSIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF030065; AAB84221.1; -.
DR HSSP; P00763; IDPO.
DR MEROPS; S01.224; -.
DR MGD; MGI:1196620; Hpn.
DR InterPro; IPR001254; -.
DR InterPro; IPR001314; -.
DR Pfam; PF00089; Trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR PROSITE; PS00134; TRYPSIN_HIS; 1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
DR KEGG; K01135; TRYPSIN; Serine protease; Transmembrane; Signal-anchor.
DR CHAIN; 1; 161; NON-CATALYTIC CHAIN (POTENTIAL).
DR CHAIN; 162; 416; CATALYTIC CHAIN (POTENTIAL).
DR DOMAIN; 1; 16; CYTOPLASMIC (POTENTIAL).
DR TRANSMEM; 17; 43; SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
(POTENTIAL).
DR FT 1; 44; 416; EXTRACELLULAR (POTENTIAL).
DR FT 162; 416; CATALYTIC.
DR FT ACT_SITE; 202; 202; CHARGE RELAY SYSTEM (BY SIMILARITY).
DR FT ACT_SITE; 256; 256; CHARGE RELAY SYSTEM (BY SIMILARITY).
DR FT ACT_SITE; 352; 352; CHARGE RELAY SYSTEM (BY SIMILARITY).
DR FT DISULFID; 152; 276; INTERCHAIN (BY SIMILARITY).
DR FT DISULFID; 187; 203; BY SIMILARITY.
DR FT DISULFID; 321; 337; BY SIMILARITY.
DR FT DISULFID; 348; 380; BY SIMILARITY.
DR FT CARBOHYD; 111; 111; N-LINKED (GLCNAC. . .) (POTENTIAL).
DR SEQUENCE; 416 AA; 44739 MW; 432194FF4004F848 CRC64;

```

Query Match 38.5%; Score 507.5; DB 1; Length 416;
 Best Local Similarity 40.4%; Pred. No. 1.6e-38;
 Matches 99; Conservative 42; Mismatches 93; Indels 11; Gaps 3;

```

QY 1 RIVGESALPGAMQVSLVNVVHVGCGSIIRPEIVTAHCVKRLNPMWHTAFAGI 60
| | | | | : | | | | | : | | | | | : | | | | | : | | | | |
DB 161 RIVGGDSSISGRPMQVSLRYDGTSLHCGSLSGDWVLTAAHCFPEKRLSKRYVAGA 220
| | | | | : | | | | | : | | | | | : | | | | | : | | | | |
QY 61 LROSEFMFGAGYOVERKVISHPNT-----DSKTRNDNDIALMKLQKPLTFNDLVKPVCLPN 114

```

```

RESULT 12
ENTR_MOUSE STANDARD: PRT; 1069 AA.
ID ENTR_MOUSE
AC P97435;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE ENTEROPEPTIDASE (EC 3.4.21.9) (ENTEROKINASE).
GN PRS7 OR ENTR.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6; TISSUE=Duodenum;
RX MEDLINE=98147142; PubMed=9486188;
RX Yuan X., Zheng X., Lu D., Rubin D.C., Pung C.Y.M., Sadler J.E.;
RT Structure of murine enterokinase (enteropeptidase) and expression in
RT small intestine during development.
RT Am. J. Physiol. 274:G342-G349(1998).
CC - FUNCTION: RESPONSIBLE FOR INITIATING ACTIVATION OF PANCREATIC
CC PROTEOLYTIC PROENZYMES (TRYPSIN, CHYMOTRYPSIN AND CARBOXYPEPTIDASE
CC A). IT CATALYZES THE CONVERSION OF TRYPSINOGEN TO TRYPSIN WHICH IN
CC TURN ACTIVATES OTHER PROENZYMES INCLUDING CHYMOTRYPSINOGEN,
CC PROCARBOXYPEPTIDASES, AND PROELASTASES (BY SIMILARITY).
CC - CATALYTIC ACTIVITY: SELECTIVE CLEAVAGE OF 6-LYS-1-ILE-7 BOND IN
CC TRYPSINOGEN.
CC - SUBUNIT: HETERODIMER OF A CATALYTIC (LIGHT) CHAIN AND A
CC MULTIDOMAIN (HEAVY) CHAIN LINKED BY A DISULFIDE BOND (BY
CC SIMILARITY).
CC - SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN (PROBABLE).
CC - PTM: THE CHAINS ARE DERIVED FROM A SINGLE PRECURSOR THAT IS
CC CLEAVED BY A TRYPSIN-LIKE PROTEASE (BY SIMILARITY).
CC - SIMILARITY: CONTAINS 2 LDL-RECEPTOR CLASS A DOMAINS.
CC - SIMILARITY: CONTAINS 2 CUB DOMAINS.
CC - SIMILARITY: CONTAINS 1 SRCR DOMAIN.
CC - SIMILARITY: CONTAINS 1 MAM DOMAIN.
CC - SIMILARITY: CONTAINS 1 MAM DOMAIN.
CC - SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC TRYPSIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U73378; AAB37317.1; -.
DR HSSP; P00763; IDPO.
DR MEROPS; S01.156; -.
DR MGD; MGI:1197523; Prs7.
DR InterPro; IPR000082; -.
DR InterPro; IPR000859; -.
DR InterPro; IPR000988; -.
DR InterPro; IPR001190; -.

```

```

DR InterPro: IPR001254;
DR InterPro: IPR001314;
DR InterPro: IPR002172;
DR Pfam: PF00431; CUB; 2.
DR Pfam: PF00629; MAM; 1.
DR Pfam: PF01390; SER; 1.
DR Pfam: PF00530; SCR; 1.
DR Pfam: PF00057; IGL_recept.a; 2.
DR Pfam: PF00089; trypsin; 1.
DR PRINTS: PR00722; CHYMOTRYPSIN.
DR PROSITE: PS00134; TRYPSIN_HIS; 1.
DR PROSITE: PS00135; TRYPSIN_SER; 1.
DR PROSITE: PS01180; CUB; 2.
DR PROSITE: PS00740; MAM; 1.
DR PROSITE: PS50060; MAM; 2.
DR PROSITE: PS01209; LDLRA; 1.
DR PROSITE: PS50068; LDLRA; 2.
KW Signal-anchor: Glycoprotein; Myristate; Hydrolase;
KW Serine protease; zymogen; Transmembrane; Repeat.
FT CHAIN 1 829 NON-CATALYTIC CHAIN (HEAVY CHAIN).
FT TRANSMEM 19 47 CATALYTIC CHAIN (LIGHT CHAIN).
FT DOMAIN 227 268 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN).
FT DOMAIN 270 379 LDL-RECEPTOR CLASS A.1.
FT DOMAIN 387 549 CUB.
FT DOMAIN 569 679 MAM.
FT DOMAIN 686 724 LDL-RECEPTOR CLASS A.2.
FT DOMAIN 723 816 SRCR.
FT ACT_SITE 874 874 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 925 925 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 1021 1021 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT LITPD 2 MYRISTATE (POTENTIAL).
FT DISULFID 229 242 BY SIMILARITY.
FT DISULFID 236 255 BY SIMILARITY.
FT DISULFID 249 266 BY SIMILARITY.
FT DISULFID 688 700 BY SIMILARITY.
FT DISULFID 695 713 BY SIMILARITY.
FT DISULFID 707 722 BY SIMILARITY.
FT DISULFID 817 945 INTERCHAIN (BY SIMILARITY).
FT DISULFID 859 875 BY SIMILARITY.
FT DISULFID 959 1027 BY SIMILARITY.
FT DISULFID 991 1006 BY SIMILARITY.
FT DISULFID 1017 1045 BY SIMILARITY.
FT CARBOHYD 147 147 N-LINKED (GLCNAG. . .) (POTENTIAL).
FT CARBOHYD 197 197 N-LINKED (GLCNAG. . .) (POTENTIAL).
FT CARBOHYD 212 212 N-LINKED (GLCNAG. . .) (POTENTIAL).
FT CARBOHYD 373 373 N-LINKED (GLCNAG. . .) (POTENTIAL).
FT CARBOHYD 380 380 N-LINKED (GLCNAG. . .) (POTENTIAL).
FT CARBOHYD 433 433 N-LINKED (GLCNAG. . .) (POTENTIAL).
FT CARBOHYD 515 515 N-LINKED (GLCNAG. . .) (POTENTIAL).
FT CARBOHYD 579 579 N-LINKED (GLCNAG. . .) (POTENTIAL).
FT CARBOHYD 675 675 N-LINKED (GLCNAG. . .) (POTENTIAL).
FT CARBOHYD 727 727 N-LINKED (GLCNAG. . .) (POTENTIAL).
FT CARBOHYD 751 751 N-LINKED (GLCNAG. . .) (POTENTIAL).
FT CARBOHYD 770 770 N-LINKED (GLCNAG. . .) (POTENTIAL).
FT CARBOHYD 791 791 N-LINKED (GLCNAG. . .) (POTENTIAL).
FT CARBOHYD 897 897 N-LINKED (GLCNAG. . .) (POTENTIAL).
FT CARBOHYD 936 936 N-LINKED (GLCNAG. . .) (POTENTIAL).
FT CARBOHYD 999 999 N-LINKED (GLCNAG. . .) (POTENTIAL).
SO SEQUENCE 1069 AA; 118735 MW; E62549E463743C3D CRC64;
Query Match 38.2%; Score 504; DB 1; Length 1069;
Best Local Similarity 41.2%; Pred. No. 9.3e-38;
Matches 98; Conservative 50; Mismatches 82; Indels 8; Gaps 4;
OY 1 RIVGESALPGAMQVSLHVNH-----VCGSGITTEPMITTAHCVKPLNNPMHTA 56
DB 829 RIVGSDAQAGAMPVVALYHRRDSTDRLLCGASLVSSDWLVAACHYRNLDPTRMTA 888
OY 57 FAGIGROSFMEFYAGYQ--VEKVISHPNVDSKTKNNDIAMLKQPLTFNDVLRKPVCLPN 114
DB 889 VGLGLMQSNLSPQVYRVRVDQIVINPHYDRRKRYNDIAMMLERKVVNTDITQICLPE 948

```

```

OY 115 PGMMLPEQLCMISGWATE-EKGTSEVLNNAKYLLETORCNSRYVNDLITPAMICA 173
DB 949 ENQIFIPGRTCISLACGYKINAGSVVDVLEKADYPLISNEKCCQQLPEYN-ITSMICA 1007
OY 174 GFLQGVNDSGCGDGGPLVTSKNNIMWLIGDTSGSGCAKATRPQYGNVMTWIT 231
DB 1008 GYEEGIDSCQDSSGPILMCOENNRFLVGVTSFGVOCALPNHPGVYRVVSQFIEMH 1065

RESULT 13
PLMN_MOUSE STANDARD: PRT: 812 AA.
ID PLMN_MOUSE
AC P20918;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE PLASMINOGEN PRECURSOR (EC 3.4.21.7) [CONTAINS: ANGIOSTATIN].
GN PLG.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-91184812; PubMed=2081600;
RA Degen S.J., Bell S.M., Schaefer L.A., Elliott R.W.;
RT "Characterization of the cDNA coding for mouse plasminogen and
RL localization of the gene to mouse chromosome 17.";
RN Genomics 8:49-61(1990).
RP CHARACTERIZATION OF ANGIOSTATIN, AND PARTIAL SEQUENCE.
RX MEDLINE-95042728; PubMed=7525077;
RA O'Reilly M.S., Holmgren L., Shing Y., Chen C., Rosenthal R.A.,
RA Moses M., Lane W.S., Cao Y., Sage E.H., Folkman J.;
RT "Angiostatin: a novel angiogenesis inhibitor that mediates the
RL suppression of metastases by a Lewis lung carcinoma.";
RN Cell 79:315-328(1994).
RP -1- FUNCTION: PLASMIN DISSOLVES THE FIBRIN OF BLOOD CLOTS AND ACTS AS
A PROTEOLYTIC FACTOR IN A VARIETY OF OTHER PROCESSES INCLUDING
EMBRYONIC DEVELOPMENT, TISSUE REMODELING, TUMOR INVASION,
AND INFLAMMATION. IN OVULATION IT WEAKENS THE WALLS OF THE
GRAAFIAN FOLLICLE. IT ACTIVATES THE UKONINASE-TYPE PLASMINOGEN
ACTIVATOR, COLLAGENASES AND SEVERAL COMPLEMENT ZYMOGENS, SUCH
AS C1 AND C5. IT CLEAVES FIBRIN, FIBRONECTIN, THROMBOSPONDIN,
LAMININ AND VON WILLEBRAND FACTOR.
RP -1- FUNCTION: ANGIOSTATIN IS AN ANGIOGENESIS INHIBITOR THAT BLOCKS
NEOVASCULARIZATION AND GROWTH OF EXPERIMENTAL PRIMARY AND
METASTATIC TUMORS IN VIVO.
CC -1- ENZYME REGULATION: CONVERTED INTO PLASMIN BY PLASMINOGEN
ACTIVATORS, BOTH PLASMINOGEN AND ITS ACTIVATOR BEING BOUND TO
FIBRIN. CANNOT BE ACTIVATED WITH STEPTOKINASE.
CC -1- MISCELLANEOUS: PLASMIN IS INACTIVATED BY ALPHA-2-ANTIPLASMIN
IMMEDIATELY AFTER DISSOCIATION FROM THE CLOT.
CC -1- MISCELLANEOUS: IN THE PRESENCE OF THE INHIBITOR, THE ACTIVATION
INVOLVES ONLY CLEAVAGE AFTER ARG-581, RESULTING IN 2 CHAINS HELD
TOGETHER BY 2 DISULFIDE BONDS. WITHOUT THE INHIBITOR, THE
ACTIVATION INVOLVES ALSO REMOVAL OF THE ACTIVATION PEPTIDE.
CC -1- SIMILARITY: CONTAINS 5 KRINGLE REGIONS.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1, ALSO KNOWN AS THE
TRYPSIN FAMILY. PLASMINOGEN SUBFAMILY.

-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see http://www.isb-sib.ch/announce/
or send an email to license@sib-sib.ch).
-----
EMBL: J04766; AAA50168.1;
DB PIR: A38514; A38514.
HSPB: P00747; 1PKK.

```

DR	MEROPS	S01.233	-	-	
DR	MGD	MGI.97620	plg.		
DR	InterPro:	IPR000001	-		
DR	InterPro:	IPR001254	-		
DR	InterPro:	IPR001314	-		
DR	InterPro:	IPR003014	-		
DR	Pfam:	PF000024	PAN: 1.		
DR	Pfam:	PF00051	kringle; 5.		
DR	Pfam:	PF00089	trypsin; 1.		
DR	PRINTS:	PR00018	KRINGLE.		
DR	PRINTS:	PR00722	CHYOTRYPSIN.		
DR	PROSITE:	PS00021	KRINGLE_1; 4.		
DR	PROSITE:	PS00134	TRYPSIN_HIS; 1.		
DR	PROSITE:	PS00135	TRYPSIN_SER; 1.		
DR	PROSITE:	PS00707	KRINGLE_2; 5.		
KW	Hydrolase:	Serine protease; Plasma;	Glycoprotein; Fibrinolysis;		
KW	Tissue remodeling:	Blood coagulation; Kringle;	zymogen; Signal.		
FT	SIGNAL	1	19		
FT	CHAIN	20	812		PLASMINOGEN.
FT	CHAIN	20	581		PLASMIN HEAVY CHAIN A.
FT	PEPTIDE	20	581		ACTIVATION PEPTIDE.
FT	CHAIN	98	561		PLASMIN SHORT FORM OF CHAIN A.
FT	CHAIN	98	7436		ANGIOSTATIN.
FT	CHAIN	582	812		PLASMIN LIGHT CHAIN B.
FT	DOMAIN	103	181		KRINGLE 1.
FT	DOMAIN	184	262		KRINGLE 2.
FT	DOMAIN	275	352		KRINGLE 3.
FT	DOMAIN	377	454		KRINGLE 4.
FT	DOMAIN	481	560		KRINGLE 5.
FT	DOMAIN	582	812		CATALYTIC.
FT	ACT_SITE	624	624		CHARGE RELAY SYSTEM (BY SIMILARITY)
FT	ACT_SITE	667	667		CHARGE RELAY SYSTEM (BY SIMILARITY)
FT	ACT_SITE	762	762		CHARGE RELAY SYSTEM (BY SIMILARITY)
FT	DISULFID	49	73		BY SIMILARITY.
FT	DISULFID	53	61		BY SIMILARITY.
FT	DISULFID	103	181		BY SIMILARITY.
FT	DISULFID	124	164		BY SIMILARITY.
FT	DISULFID	152	176		BY SIMILARITY.
FT	DISULFID	185	262		BY SIMILARITY.
FT	DISULFID	188	316		BY SIMILARITY.
FT	DISULFID	206	245		BY SIMILARITY.
FT	DISULFID	234	257		BY SIMILARITY.
FT	DISULFID	275	352		BY SIMILARITY.
FT	DISULFID	296	335		BY SIMILARITY.
FT	DISULFID	324	347		BY SIMILARITY.
FT	DISULFID	377	454		BY SIMILARITY.
FT	DISULFID	398	437		BY SIMILARITY.
FT	DISULFID	426	449		BY SIMILARITY.
FT	DISULFID	481	560		BY SIMILARITY.
FT	DISULFID	502	543		BY SIMILARITY.
FT	DISULFID	531	555		BY SIMILARITY.
FT	DISULFID	568	687		INTERCHAIN (BY SIMILARITY).
FT	DISULFID	578	586		INTERCHAIN (BY SIMILARITY).
FT	DISULFID	609	625		BY SIMILARITY.
FT	DISULFID	701	768		BY SIMILARITY.
FT	DISULFID	731	747		BY SIMILARITY.
FT	DISULFID	758	766		BY SIMILARITY.
SO	SEQUENCE	812 AA;	90846 MW;	D3A47AA4FC2256F8 CRC64;	

```

Query Match 38.1%; Score 502.5; DB 1; Length 812:
Best Local Similarity 43.3%; Pred. No. 9,4e-38;
Matches 106; Conservative 36; Mismatches 78; Indels 25; Gaps 7;

Oy 1 RIVGESALPGAMPQVSLHVO--NVHVGSGSIIPPEVITAAHCYERKLPNNMHTATA 58
   |||| | : |||| | : | : ||| : ||| : |||| | : | :
Db 561 RYVGCGVAMPHPMQIISLRTFTGQHCGLIPAEWVITAAHCLERS-SRPEFYKYL 639
   |||| | : |||| | : | : ||| : ||| : |||| | : | :
Oy 59 GILRSFMFYGAGYQVE-----KVISHPNYSKTNNDIALMLKQPLTFNDLVKPVCLP 113
   |||| | : |||| | : | : ||| : ||| : |||| | : | :
Db 640 GAHEE-----YINGLDVQELSVAKLLPER-----NRDIALKLKSRPATITQKVIAPCLP 689
   |||| | : |||| | : | : ||| : ||| : |||| | : | :
Oy 114 NFGMLQBPQLCWMISGWATE--EKGRTSEVLANAKVLLIETGRNSRYYVDNLTTPAM 170

```

```

Db      690 SPNVYVADRTICITCYTGMKETOGETFCAGALKE-...AQLPVLENKACNRYELNNRVKSTE 745
QY      171 ICAGFLQGNVDSQCGSDSGPLVYTSKNKNIMWMLIGDTSMSGCAKAYRPGVNGVWVFTDMI 230
Db      746 LCAGGLAGGVDSQCGSDSGPLVCFEKKDXIILQGVYTWGLGCAKRPKPKGVYRVSRVFDVMI 805
QY      231 YRQMR 235
Db      806 BREMR 810

RESULT 14
ST14_HUMAN STANDARD: PRT; 855 AA.
ID ST14_HUMAN
AC Q9Y5Y6;
DT 01-OCT-2000 (Rel. 40, Created)
DT 01-OCT-2000 (Rel. 40, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE SUBSEQUENCE OF TUMORICENICITY 14 (EC 3.4.21.-) (MATRIPTASE) (MEMBRANE-
DE TYPE SERINE PROTEASE 1) (MT-SP1).
DE ST14 OR PRSS14 OR SNCL9.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99303581; PubMed=10373424;
RA Lin C.Y., Anders J., Johnson M., Sang Q.A., Dickson R.B.;
RT "Molecular cloning of cDNA for matriptase, a matrix-degrading serine
RT protease with trypsin-like activity."
RN J. Biol. Chem. 274:18231-18236(1999).
RN [2]
RP SEQUENCE FROM N.A.
RA Takeuchi T., Shuman M.A., Craik C.S.;
RT "Reverse biochemistry: Use of macromolecular protease inhibitors to
RT dissect complex biological processes and identify a membrane-type
RT serine protease in epithelial cancer and normal tissue."
RN Proc. Natl. Acad. Sci. U.S.A. 96:11054-11061(1999).
RN [3]
RP CHARACTERIZATION.
RX TISSUE=Milk;
RX PubMed=10373425;
RX Lin C.Y., Anders J., Johnson M., Dickson R.B.;
RT "Purification and characterization of a complex containing matriptase
RT and a Kunitz-type serine protease inhibitor from human milk."
RN J. Biol. Chem. 274:18237-18242(1999).
CC -1- FUNCTION: DEREGATES EXTRACELLULAR MATRIX. PROPOSED TO PLAY A ROLE
CC IN BREAST CANCER INVASION AND METASTASIS. EXHIBITS TRYPSIN-LIKE
CC ACTIVITY AS DEFINED BY CLEAVAGE OF SYNTHETIC SUBSTRATES WITH ARG
CC OR LYS AS THE P1 SITE.
CC -1- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN (PROBABLE).
CC -1- SIMILARITY: CONTAINS 4 LBD-RECEPTOR CLASS 4 DOMAINS.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC TRYPSIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL, AF118224; AAD42765.2; -
DR EMBL, AF133086; AAF00109.1; -
DR HSSP: P00763; IDPO.
DR MEROPS: S01.302; -
DR InterPro: IPR000859; -
DR InterPro: IPR001254; -
DR InterPro: IPR001314; -

```

```

DR InterPro: IPR002172; -.
DR Pfam: PF00057; 1dL_recept.a; 4.
DR Pfam: PF00089; trypsin; 1.
DR Pfam: PF00431; CUB; 2.
DR PRINTS: PR00261; LDLRECEPTOR.
DR PRINTS: PR00722; CHYMOTRYPSIN.
DR PROSITE: PS00134; TRYPSIN_HIS; 1.
DR PROSITE: PS00135; TRYPSIN_SER; 1.
DR PROSITE: PS01180; CUB; 2.
DR PROSITE: PS01209; LDLRA.1; 2.
DR PROSITE: PS50068; LDLRA.2; 4.
DR Signal-anchor: Glycoprotein; Hydrolyase; Serine protease;
KM Transmembrane; Repeat.
KM TRANSMEM 1 55
FT DOMAIN 56 76
FT DOMAIN 77 855
FT DOMAIN 214 334
FT DOMAIN 340 447
FT DOMAIN 452 487
FT DOMAIN 487 524
FT DOMAIN 524 560
FT DOMAIN 566 603
FT DOMAIN 614 851
FT ACT_SITE 656 656
FT ACT_SITE 711 711
FT ACT_SITE 805 805
FT CARBOHYD 109 109
FT CARBOHYD 302 302
FT CARBOHYD 485 485
FT CARBOHYD 772 772
FT SEQUENCE 855 AA; 94769 MW; 26143132C01P99C9 CRC64;

Query Match 38.1%; Score 502; DB 1; Length 855;
Best Local Similarity 41.7%; Pred. No. 1,1e-37;
Matches 100; Conservative 45; Mismatches 81; Indels 14; Gaps 7;

QY 1 RIVGESALPGAMPQVSLH-VQNVHVGSGSIITPEWITAAHCVEKPE---LNNPMHWT 55
DB 614 RIVGGTDADEGEMPMQVSLHVLGQGHICGASLISPMVLVSAHNCYIDRGRFSDPTQWT 673
QY 56 AFACILRGSEFMVYGAGYQ---VEKVISHPNYSKTKKNDIALMKLOKPLFNDLYKPCVL 112
DB 674 AFLGLHDS-QRSAPGVERRKRKRITISHPEFNDFFDYDIALELEKREKYSVMRPTCL 732
QY 113 PNPGMMLQPEQLCWMISGMGATEEKGKTSSEVLNAKVLLEIQRCSRYVDNLITPMAMIC 172
DB 733 PDASHVFPAGAKAIWVTGMGHQYGTGALLLOKGEIRVINQTTCE--LLPQOITPRMWC 790
QY 173 AGFLQGNVDSGCGSGGPL--VTSKNNTIMWLIGDTSWGSCKAKAYRPGYGVNVVFTDMI 230
DB 791 VGFISGVDSCGSGGGLSSVEADGRI-FOAGVYVWDGCAQRNKPQVYTRLPLEFRDI 849

RESULT 15
HEPS_RAT STANDARD: PRT; 416 AA.
AC Q05511;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE SERINE PROTEASE HEPSIN (EC 3.4.21.-).
GN HPN.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_Taxid=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=9330573; PubMed=8318546;
RA Farley D., Raymond F., Nick H.;

```

```

RT "Cloning and sequence analysis of rat hepsin, a cell surface serine
RT proteinase.";
RL Biochim. Biophys. Acta 1173:350-352(1993).
CC -1- FUNCTION: PLAYS AN ESSENTIAL ROLE IN CELL GROWTH AND MAINTENANCE
CC OF CELL MORPHOLOGY.
CC -1- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC TRYPSIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X70900; CA50256.1; -.
DR PIR: S32013; S32013.
DR PIR: S33777; S33777.
DR HSRP: P00763; IDPO.
DR MEROPS: S01.224; -.
DR InterPro: IPR001254; -.
DR InterPro: IPR001314; -.
DR Pfam: PF00089; trypsin; 1.
DR PRINTS: PR00722; CHYMOTRYPSIN.
DR PROSITE: PS00134; TRYPSIN_HIS; 1.
DR PROSITE: PS00135; TRYPSIN_SER; 1.
KM Hydrolyase; Serine protease; Transmembrane; Signal-anchor.
FT CHAIN 1 161
FT CHAIN 162 416
FT DOMAIN 1 16
FT DOMAIN 17 43
FT TRANSMEM 17 43
FT DOMAIN 44 416
FT DOMAIN 162 416
FT ACT_SITE 202 202
FT ACT_SITE 256 256
FT ACT_SITE 352 352
FT DISULFID 152 276
FT DISULFID 187 203
FT DISULFID 321 337
FT DISULFID 348 380
FT CARBOHYD 111 111
FT SEQUENCE 416 AA; 44926 MW; ESA9E8FA950E180 CRC64;

Query Match 37.9%; Score 499.5; DB 1; Length 416;
Best Local Similarity 40.0%; Pred. No. 8.3e-38;
Matches 98; Conservative 43; Mismatches 55; Indels 11; Gaps 3;

QY 1 RIVGESALPGAMPQVSLH-VQNVHVGSGSIITPEWITAAHCVEKPELNNPMHWTAFAGI 60
DB 161 RIVGGDSSLGRMPQVSLRYDGTGLCGSLSDGWLITAAHCPEERRRVLSSRVFAGA 220
QY 61 LRQSPFYGAGQVVEKVISHPNY-----DSKTKNDIALMKLOKPLFNDLYKPCVLN 114
DB 221 VARTSP-HAVOLGQAVYVYHGGYLPFRDPTIDENSNDIALVLLSSLPLETEIQVCLPA 279
QY 115 PNPMMLQPEQLCWMISGMGATEEKGKTSSEVLNAKVLLEIQRCSRYVDNLITPMAMICAG 174
DB 280 AGQALVDGKVCYVTGMGNTQFYGGQAVVLOEAKRYVITISNEVCNSDFTGNOIKPKMFCAG 339
QY 175 FLQGNVDSGCGSGGPL--VTSKNNTIMWLIGDTSWGSCKAKAYRPGYGVNVVFTDMI 230
DB 340 YPEGIDACGCGSGGHFVCEIDRISGTSRMRLCGIYSWGTGCLARKPQVYTRKVIDFERMI 399
QY 231 YKQMR 235
DB 400 FOAIR 404

```

Wed Sep 26 13:57:02 2001

us-09-615-285-2_copy_255_492.rsp

Page 15

Search completed: September 26, 2001, 09:27:27
Job time: 317 sec

THIS PAGE BLANK (USPTO)

Qy	181	DSOGBDSGGGLYTSKNNIMWLIGDTSMSGGCKAKATRPVYGVNAYVFDMIRQMR	235
	1		111
	2		111
	3		111
	4		111
	5		111
	6		111
	7		111
	8		111
	9		111
	10		111
	11		111
	12		111
	13		111
	14		111
	15		111
	16		111
	17		111
	18		111
	19		111
	20		111
	21		111
	22		111
	23		111
	24		111
	25		111
	26		111
	27		111
	28		111
	29		111
	30		111
	31		111
	32		111
	33		111
	34		111
	35		111
	36		111
	37		111
	38		111
	39		111
	40		111
	41		111
	42		111
	43		111
	44		111
	45		111
	46		111
	47		111
	48		111
	49		111
	50		111
	51		111
	52		111
	53		111
	54		111
	55		111
	56		111
	57		111
	58		111
	59		111
	60		111
	61		111
	62		111
	63		111
	64		111
	65		111
	66		111
	67		111
	68		111
	69		111
	70		111
	71	DTCCGDSGGGLYVNRKGTWMLVGDFTSMGGCCARANKPCVYGVNAYVFLEWISQMR	764
RESULT	2		
Q9H3S3			
ID	Q9H3S3	PRELIMINARY;	PRT; 457 AA.
AC	Q9H3S3;		
DT	01-MAR-2001 (TrEMBLrel. 16, Created)		
DT	01-MAR-2001 (TrEMBLrel. 16, Last sequence update)		
DT	01-MAR-2001 (TrEMBLrel. 16, Last annotation update)		
DE	SPINESIN.		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.		
OX	NCBI_TaxID=9606;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	TISSUE=BRAIN PROSTATE;		
RA	Mitsui S., Yamaguchi N.;		
RL	"Molecular cloning of human spinosin.";		
DR	Submitted (MAY-1989) to the EMBL/Genbank/DBJ databases.		
DR	EMBL; AB028140; BAB20375.1;		
DR	SEQUENCE 457 AA; 49574 MW; 64406AB4985A2651 CRC64;		

Query Match	42.38	Score 558	DB 4	Length 457
Best Local Similarly	45.18	Pred. No. 2.3e-48		
Matches 105	Conservative 31	Mismatches 95	Indels 2	Gaps 2

```

Oy 1 RGGSEALGCAAPMQVSLHQNVHVCSSITTPRMTATACHEK -PLNNPHHTAFAP
Db 217 RIVGGGVADGRMPQASVALGRRHFCGSGVLAIPWVVTAAHCHSRLLRASSMRVHAG
Oy 60 ILROSEMFYAGYQVEKVIASHPNYDSKTKNDIALMLQKLELPFNDDLVKVCCLPNFGML
Db 277 LVSHASNRRPHQGLAVERIRIPHPYLSAQNHDDYVALLRLQTLNFSDFVAGVCLPAKEQHE
Oy 120 QPELCLWISGNGATEKGR-TSEVLNNAKVVLLIEFRNSKRYVDNLITPAMICAEFLQCG
Db 337 PKGRCWVSGWGHTRPSHTYSSDMIDQTVYLPFSTQLCNSSCVSGALTPTMLCAGLQGG
Oy 179 NVDSGCGDSGGLPVTSKNNIMWLGDTSMGSGCAKAPGVYGVNWFTDMY 231
Db 397 RADACGDSGGLPVCDDGTWRVLGVSWGRACAPNHPGVYAKVAEFLDMIH 449

RESULT 3
097506 PRELIMINARY; PRT; 643 AA.
AC 097506;
DT 01-MAY-1999 (TREMBLrel. 10, Created)
DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
DE KALLIKREIN.
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
ON NCBI_TaxID=9823;
[1]
RP SEQUENCE FROM N.A.
RC TISSUE=PLASMA.
RA Takahashi T., Kimura A., Okimura H., Hamabata T.;
RT "Porcine liver plasma kallikrein."
RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.
CC - SIMILARITY: TO SERINE PROTEASES, TRYPsin FAMILY.
CC - SIMILARITY: TO CHYNOTRYPSIN SERINE PROTEASE FAMILY (Sl).
DR EMBL: AB022425. BAA37147.1. -.
DR HSP: P00766. ICHG.
DR MEROPS: S01.212. -.
DR InterPro: IPR000177. -.
DR InterPro: IPR001254. -.

```

DR InterPro: IPR001314; -.
DR InterPro: IPR003014; -.
DR Pfam: PF00024; PAN; 4.
DR Pfam: PF00089; trypsin; 1.
DR PRINTS: PR00005; APPLEDOMAIN.
DR PRINTS: PR00722; CHYMOTRYPSIN.
DR PROSITE: PS00495; APPLE; 4.
DR PROSITE: PS00134; TRYPSIN_HIS; UNKNOWN_1.
DR PROSITE: PS00135; TRYPSIN_SER; UNKNOWN_1.
DR SMART: SM00020; Tryp_Spc; 1.
KW Hydrolase; Serine protease.
SQ SEQUENCE 643 AA; 72227 MW; AFF2933E3C3CB80A CRC64;

Query Match	42.0%	Score 554;	DB 6;	Length 643;
Best Local Similarity	44.2%;	Pred. No. 9e-48;		
Matches 107;	Conservative 40;	Mismatches 87;	Indels 8;	Gaps 5

QY 1 RIYGGESALPCAMPQVS-----LHVQNVHVGSGSITPEWIVTAACHVEKPLNPNWHHTA 56

01 398 KLVGGTDFLEWPMQVSLQAKRAQN-HLGGSLGHQWELIAHCFD-GSLSPDIWRI 455

```

Db      456 YGGLINSEITKETPFQVAEIIHONKYLESGHDALLKLETPNTYDFQKICLPSR 515
      : || | : ||::| || :||::|| : | ||::|

```

QY 116 GMLQPEOLCWSISGATEEKEKCTSEVLNAKVLLETQRCNSRYVDNLITPAMICAGF 175

516 DDLNNVYTCNCWTGMGFTTEKGEIQNIIQKVNIPLYSNEECQKSY RDHKISKQMICAGY 574

```
Db      575 KEGGKDACCKSSGGPLVCKYNGIWHLVGTTSWEGCGARRQPGVTYTKVIEYMIMILEKTQ 634
```

QY 236 AD 237

Db 635 Dd 636 .

RESULT 4
Q9ER02

1D	Q9EN02	FREELIMINARY,	FRI,	2/3 AM.
AC	Q9ER02;			
DE	01-MAR-2001 (WREMB1:ra)	16	Created)	

DT	01-MAR-2001 (Tremblurel. 16, Last sequence update)
DT	01-MAR-2001 (Tremblurel. 16, Last annotation update)

05 Mus musculus (Mouse).
06
07 Eukaryota: Metazoa: Chordata: Vertebrata: Euteleostomi:
08

NCBI_TaxID=10090; Mammalia; Eulnidae; Rodentia; Sciurognathii; Muridae; Mus.

RP	SEQUENCE FROM N.A.
RC	TISSUE=BRAIN;

RT "cDNA cloning of mouse type 1 splinesin.,"
Submitted (VIII-1998) to the FMI/GenBank/DBI databases

DR EMBL; AB016423; BAB202/8-1; -.
SQ SEQUENCE 273 AA; 29662 MW; 63E45FAD2BDF7F5D CRC64;

Query Match 41.68; Score 549; DB 11; Length 273;

best local similarity 42.36; Freq. NO. 5.36-46;
Matches 100; Conservative 37; Mismatches 94; Indels 2; Gaps

```

QY 1 RIVGESALPGAMPWQVSLHVQNVHVCSSITTPENIVTAHCEK-PLNNPWHMTAFAG 59
      ||||: | ||| : : | ||:: | :|||: | :

```

0x 60 IIR0SFMEVGCAGY0EKVISHPNYDSKTKNNIDIALMKLOKPILENDIYKRPVCLPNPGMML 119

```

Db      95 LVSHGAVRQHQTWEKIPHPRLYSAGNHDDYDVALQLRTPINFSDPVDVAVCTPAKEQFF 154
      ::  :  |  |||  |  ::  :  |::|::|  :  |  |  |||

```


OY 113 PNPGLMLOPELQWISGCGATEBKGTSEVLNAKVLLEIETORCSRYYVNDLITPAMIC 172
 ID 113 PNPGLMLOPELQWISGCGATEBKGTSEVLNAKVLLEIETORCSRYYVNDLITPAMIC 172
 DB 448 PSKGDNRNVTYTCWYMGWRKLRKIQNTLQAKPLVTNECCORRY-RGHKITHKMIC 506
 OY 173 AGELQGNVSCGDSGGPLVTSKNNIMWLIGDTSWGSCKAKARPGVYGNVFTDWIR 232
 DB 507 AGYREGGKADCKGDSGGPLSKHNEVWLHGLTSMGSCAQHREPRGVYVNVVEYDWILE 566
 OY 233 QMRA 236
 DB 567 KMQA 570

RESULT 8
 O9QY29 PRELIMINARY: PRT: 310 AA.
 ID 09QY29 PRELIMINARY: PRT: 310 AA.
 DB 01-MAY-2000 (TREMBLrel. 13, Created)
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
 DE 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
 DE DISTAL INTESTINAL SERINE PROTEASE.
 GN DISP.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Shaw-Smith C.J., Coffey A.J., Leversha M., Freeman T.C., Bentley D.R.,
 RT Walters J.R.:
 RT "Characterization of a novel murine intestinal serine protease,
 RT DISP."
 RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
 CC -1- SIMILARITY: TO CHYMOTRYPSIN SERINE PROTEASE FAMILY (S1).
 DB EMBL: AJ243866; CAB56465.1; -.
 DR HSSP: P20231; IANO.
 DR InterPro: IPR000038; -.
 DR InterPro: IPR001354; -.
 DR InterPro: IPR001314; -.
 DR Pfam: PF00089; trypsin; 1.
 DR PRINTS: PR00722; CHYMOTRYPSIN.
 DR PRODOM: PD002565; -; 1.
 DR PROSITE: PS00134; TRYPSIN_HIS; UNKNOWN_1.
 DR PROSITE: PS00135; TRYPSIN_SER; 1.
 DR SMART: SM00020; tryp_spec; 1.
 KW Protease.
 SO SEQUENCE 310 AA; 33701 MW; F828EC7F6D25303F CRC64;

Query Match 39.8%; Score 524.5; DB 11; Length 310;
 Best Local Similarity 42.3%; Pred. No. 3.5e-45;
 Matches 102; Conservative 41; Mismatches 87; Indels 11; Gaps 5;
 OY 1 RIVGESALPGAMPQVSLHV-QNVHVGSGSITTEPMITVAACVKEKPLNNPMHTAFAG 59
 DB 36 KIVGGQDLBEGMPQVSLHV-QNVHVGSGSITTEPMITVAACVKEKPLNNPMHTAFAG 95
 OY 60 ILKOSFMEYAGAYQVEKVI-SHPYDSKTKNDIALMLKQLPFTNDLVKPVCLPNPGM 118
 DB 96 LTLSLLEPHSTLVAVRNFVHPTLVLAADASSGDIALVQDPLRPSQFT-RVCLPRAQTF 154
 OY 119 LQPEQLCHISGKATBEKGTSEVLNAKVLLEIETORCSRYYVNDLITPAMIC 171
 DB 155 LTPGTCVAVTGCATQERDMS-VLQELAVPLDSEDEKMTHTGSSLSGERIIOQMDL 213
 OY 172 CAGFLQGNVSCGDSGGPLVTSKNNIMWLIGDTSWGSCKAKARPGVYGNVFTDWIR 231
 DB 214 CAGVEGHIDSCGDSGGPLVTSKNNIMWLIGDTSWGSCKAKARPGVYGNVFTDWIR 273
 OY 232 R 232
 DB 274 R 274

RESULT 9
 O9NZAS PRELIMINARY: PRT: 423 AA.
 ID 09NZAS PRELIMINARY: PRT: 423 AA.
 DB 01-OCT-2000 (TREMBLrel. 15, Created)
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
 DE 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
 DE TYPE II MEMBRANE SERINE PROTEASE.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Smeekens S.S., Lorimer D.D., Wang E., Hou J., Linnevers C.:
 RT "MT-SP2, a novel type II membrane serine protease expressed in
 RT trachea, colon, and small intestine: Identification, cloning, and
 RT chromosomal localization."
 RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
 CC -1- SIMILARITY: TO CHYMOTRYPSIN SERINE PROTEASE FAMILY (S1).
 DB EMBL: AF216312; AAF31436.1; -.
 DR InterPro: IPR001254; -.
 DR InterPro: IPR001314; -.
 DR InterPro: IPR002172; -.
 DR Pfam: PF00057; LDL_recept_a; 1.
 DR Pfam: PF00089; trypsin; 1.
 DR PRINTS: PR00722; CHYMOTRYPSIN.
 DR PROSITE: PS00134; TRYPSIN_HIS; UNKNOWN_1.
 DR PROSITE: PS00135; TRYPSIN_SER; 1.
 DR SMART: SM00192; LDLa; 1.
 KW Protease.
 SO SEQUENCE 423 AA; 46397 MW; 90792AF0F8AFE9A30 CRC64;

Query Match 39.5%; Score 521; DB 4; Length 423;
 Best Local Similarity 44.6%; Pred. No. 1.2e-44;
 Matches 108; Conservative 36; Mismatches 84; Indels 14; Gaps 7;
 OY 1 RIVGESALPGAMPQVSLHV-QNVHVGSGSITTEPMITVAACVKEKPLNNPMHTAFAG 60
 DB 190 RYVGEAEASVDSMPQVSLHV-QNVHVGSGSITTEPMITVAACVKEKPLNNPMHTAFAG 248
 OY 61 LR-QSFMEYAGAYQVEKVI-SHPYDSKTKNDIALMLKQLPFTNDLVKPVCLPNPG 116
 DB 249 DKLSGF---PSLAVALKIIIEFNPMY---PRNDIALMLKQLPFTSGSTVAPICLPFFD 301
 OY 117 MMLQPEQLCHISGKATBEKGTSEVLNAKVLLEIETORCSRYYVNDLITPAMIC 175
 DB 302 EELTPATPLMIIIGMGFTKONGKMSDILLQASVQYIDSTRCADADAYQGEVEKMKCAGI 361
 OY 176 LOGNVSCGDSGGPLVTSKNNIMWLIGDTSWGSCKAKARPGVYGNVFTDWIR 235
 DB 362 PEGVYDTCGDSGGPLVTSKNNIMWLIGDTSWGSCKAKARPGVYGNVFTDWIR 420
 OY 236 AD 237
 DB 421 AE 422
 RESULT 10
 O9NRS4 PRELIMINARY: PRT: 437 AA.
 ID 09NRS4 PRELIMINARY: PRT: 437 AA.
 DB 01-OCT-2000 (TREMBLrel. 15, Created)
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
 DE 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
 DE TRANSMEMBRANE SERINE PROTEASE 3.
 GN TMPSR3.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 OX NCBI_TaxID=9606;

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 26, 2001, 09:22:08 : Search time 20.35 Seconds
(without alignments)
240.811 Million cell updates/sec

Title: US-09-615-285-2_COPY_255_492
Perfect score: 1319
Sequence: 1 RIVGESALPGAMPWQVSLH.....VYGNVMVFTDWIYRQNRADG 238

Scoring table:
BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 197339 seqs, 20590346 residues

Total number of hits satisfying chosen parameters: 197339

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents_AA:*

- 1: /cgn2_6/ptodata/2/1aa/5A.COMB.pep:*
- 2: /cgn2_6/ptodata/2/1aa/5B.COMB.pep:*
- 3: /cgn2_6/ptodata/2/1aa/6A.COMB.pep:*
- 4: /cgn2_6/ptodata/2/1aa/6B.COMB.pep:*
- 5: /cgn2_6/ptodata/2/1aa/PCTOS.COMB.pep:*
- 6: /cgn2_6/ptodata/2/1aa/Backfile1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1319	100.0	492	4 US-09-342-749-2	Sequence 2, Appl1
2	1310	99.3	283	3 US-08-807-151-1	Sequence 1, Appl1
3	553.5	42.0	248	4 US-08-944-483-63	Sequence 63, Appl
4	543	41.2	638	2 US-08-681-151-3	Sequence 3, Appl1
5	528.5	40.1	256	2 US-09-027-337-3	Sequence 3, Appl1
6	524	39.7	238	4 US-08-944-483-64	Sequence 64, Appl
7	523.5	39.7	255	4 US-08-944-483-67	Sequence 67, Appl
8	523.5	39.7	798	4 US-08-200-900A-2	Sequence 2, Appl1
9	523.5	39.7	798	5 PCT-US94-00616-2	Sequence 2, Appl1
10	521.5	39.5	235	3 US-08-807-151-3	Sequence 3, Appl1
11	521	39.5	435	4 US-09-008-271A-6	Sequence 6, Appl1
12	518.5	39.3	418	1 US-08-508-448C-25	Sequence 25, Appl
13	513.5	38.9	232	1 US-08-508-448C-19	Sequence 19, Appl
14	511.5	38.8	235	4 US-08-944-483-65	Sequence 65, Appl
15	507.5	38.5	416	2 US-09-000-846-2	Sequence 2, Appl1
16	502.5	38.1	812	1 US-08-248-629A-1	Sequence 1, Appl1
17	502.5	38.1	812	1 US-08-451-932-1	Sequence 1, Appl1
18	502.5	38.1	812	1 US-08-452-260-1	Sequence 1, Appl1
19	502.5	38.1	812	1 US-08-326-785-1	Sequence 1, Appl1
20	502.5	38.1	812	2 US-08-612-788-1	Sequence 1, Appl1
21	502.5	38.1	812	2 US-08-605-598B-1	Sequence 1, Appl1
22	502.5	38.1	812	2 US-08-429-743-1	Sequence 1, Appl1
23	502.5	38.1	812	2 US-08-866-735-1	Sequence 1, Appl1
24	502.5	38.1	812	2 US-09-066-028-1	Sequence 1, Appl1
25	502.5	38.1	812	5 PCT-US95-05107-1	Sequence 1, Appl1
26	502	38.1	855	5 US-09-027-337-2	Sequence 2, Appl1
27	495.5	37.6	250	4 US-08-944-483-68	Sequence 68, Appl

28	459	34.8	356	2 US-08-681-151-1	Sequence 1, Appl1
29	455.5	34.5	314	4 US-09-008-271A-3	Sequence 3, Appl1
30	450	34.1	270	2 US-08-978-404B-8	Sequence 8, Appl1
31	447	33.9	276	2 US-09-016-366A-15	Sequence 15, Appl1
32	447	33.9	276	2 US-08-978-404B-21	Sequence 21, Appl1
33	446	33.8	230	1 US-08-379-621-2	Sequence 2, Appl1
34	446	33.8	230	1 US-08-147-000B-2	Sequence 2, Appl1
35	446	33.8	230	2 US-08-889-078-2	Sequence 2, Appl1
36	444	33.7	299	4 US-08-944-483-66	Sequence 66, Appl
37	443	33.6	791	1 US-08-643-219-1	Sequence 1, Appl1
38	443	33.6	791	1 US-08-851-350-1	Sequence 1, Appl1
39	442.5	33.5	221	4 US-08-944-483-54	Sequence 54, Appl1
40	442.5	33.5	222	1 US-08-456-840-46	Sequence 46, Appl
41	442.5	33.5	222	1 US-08-266-407A-6	Sequence 46, Appl
42	442.5	33.5	222	2 US-08-892-544-46	Sequence 46, Appl
43	442	33.5	546	6 5200340-6	Sequence 46, Appl
44	442	33.5	790	1 US-08-469-486-54	Sequence 54, Appl
45	442	33.5	790	2 US-08-469-658-54	Sequence 54, Appl

ALIGNMENTS

RESULT 1
US-09-342-749-2
Sequence 2, Application US/09342749
Patent No. 6166194
GENERAL INFORMATION:
APPLICANT: Mong, Alexander K.C.
APPLICANT: Tavligian, Sean V.
APPLICANT: Teng, David H.-F.
APPLICANT: Wyriad Genetics, Inc.
TITLE OF INVENTION: TMPRSS2 is a Tumor Suppressor
FILE REFERENCE: 2318-202
CURRENT APPLICATION NUMBER: US/09/342,749
CURRENT FILING DATE: 1999-06-29
EARLIER APPLICATION NUMBER: US 60/091,044
EARLIER FILING DATE: 1998-06-29
NUMBER OF SEQ ID NOS: 33
SOFTWARE: Patentln Ver. 2.0
SEQ ID NO 2
LENGTH: 492
TYPE: PRT
ORGANISM: Homo sapiens
US-09-342-749-2

Query Match 100.0%; Score 1319; DB 4; Length 492;
Best Local Similarity 100.0%; Pred. No. 8.7e-139;
Matches 238; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RIVGESALPGAMPWQVSLHVNWHVCGSIITPEWIVTAHCEVRLPNNPMTAFAGI 60
|||||
DB 255 RIVGESALPGAMPWQVSLHVNWHVCGSIITPEWIVTAHCEVRLPNNPMTAFAGI 314
QY 61 LRSFMEYAGYGVVERVISHPNVDSKTKNDIALMLQKPLTFNDLVKPCLPNPMMLQ 120
|||||
DB 315 LRSFMEYAGYGVVERVISHPNVDSKTKNDIALMLQKPLTFNDLVKPCLPNPMMLQ 374
QY 121 PEOLCISWGATEEKGKTESEVLNAKVLLETORNSRYVDNLITPAMICAGFLQGNV 180
|||||
DB 375 PEOLCISWGATEEKGKTESEVLNAKVLLETORNSRYVDNLITPAMICAGFLQGNV 434
QY 181 DSCQDSGSGPLVTSKNNIMWLIDITSGSCCAKAYRPGYGNVFTDWIYRQNRADG 238
|||||
DB 435 DSCQDSGSGPLVTSKNNIMWLIDITSGSCCAKAYRPGYGNVFTDWIYRQNRADG 492

RESULT 2
US-08-807-151-1
Sequence 1, Application US/08807151
Patent No. 6043033
GENERAL INFORMATION:

APPLICANT: Bandman, Olga
 APPLICANT: Lal, Preeti
 TITLE OF INVENTION: NOVEL HUMAN PROSTATE-ASSOCIATED
 TITLE OF INVENTION: PROTEASE
 NUMBER OF SEQUENCES: 5
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Incyte Pharmaceuticals, Inc.
 STREET: 3174 Porter Drive
 CITY: Palo Alto
 STATE: CA
 COUNTRY: US
 ZIP: 94304
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Diskette
 COMPUTER: IBM Compatible
 OPERATING SYSTEM: DOS
 SOFTWARE: FASTSEQ Version 2.0
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/807,151
 FILING DATE: Filed Herewith
 CLASSIFICATION: 424
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER:
 FILING DATE:
 ATTORNEY/AGENT INFORMATION:
 NAME: Billings, Lucy J.
 REGISTRATION NUMBER: 36,749
 REFERENCE/DOCKET NUMBER: PF-0227 US
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 415-845-0555
 TELEFAX: 415-845-4166
 INFORMATION FOR SEQ ID NO: 1:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 283 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 IMMEDIATE SOURCE:
 LIBRARY: SCORNOT01
 CLONE: 556016
 US-08-807-151-1

```

Query Match Summary          99.3%: Score 1310: DB 3: Length 283:
Best Local Similarity       99.6%: Pred. NO. 3.9e-139:
Matches 237: Conservative 0: Mismatches 1: Indels 0: Gaps 0:

OY      1  RIVGESALPGAMPQVSLHVNQNVHVCSSITTPEMITVAACHCEKPLNNPWHMTAFAGI 60
Db      46  RIVGESALPGAMPQVSLHVNQNVHVCSSITTPEMITVAACHCEKPLNNPWHMTAFAGI 105

OY      61  LRQSMFPGAGYQVKEKVIISHPNYDSKTKNDIALMKLQKPLTFENDLVKPVCLPNPGMMIQ 120
Db     106  LRQSMFPGAGYQVKEKVIISHPNYDSKTKNDIALMKLQKPLTFENDLVKPVCLPNPGMMIQ 165

OY     121  PEOLCWMISGMGATEKRGKTSFVLNAKAYLLITPQCNNSRYVVDNLITPAMICAGFLQGNV 180
Db     166  PEOLCWMISGMGATEKRGKTSFVLNAKAYLLITPQCNNSRYVVDNLITPAMICAGFLQGNV 225

OY     181  DSCQDSCGPLYTSKNNTIMWLLIGDTSMWSSGCAKARPGYGVGVWVETFDIYROMRADG 238
Db     226  DSCQDSCGPLYTSKNNTIMWLLIGDTSMWSSGCAKARPGYGVGVWVETFDIYROMRADG 283

RESULT      3
: US-08-944-483-63
: Sequence 63, Application US/08944483
: Patent No. 6232456
: GENERAL INFORMATION:
: APPLICANT: COHEN, MAURICE
: APPLICANT: COLPITTS, TRACEY L.
: APPLICANT: FRIEDMAN, PAULA N.
: APPLICANT: GRANNADOS, EDWARD N.

```

1 APPLICANT: KASS, MICHAEL R.
 2 APPLICANT: RUSSELL, JOHN C.
 3 APPLICANT: STEWART, KENT D.
 4 APPLICANT: STROUPE, STEVEN D.
 5 TITLE OF INVENTION: NOVEL SERINE PROTEASE REAGENTS
 6 TITLE OF INVENTION: AND METHODS USEFUL FOR DETECTING AND TREATING DISEASES
 7 TITLE OF INVENTION: OF THE PROSTATE
 8 NUMBER OF SEQUENCES: 76
 9 CORRESPONDENCE ADDRESS:
 10 ADDRESSEE: Abbott Laboratories
 11 STREET: 100 Abbott Park Road
 12 CITY: Abbott Park
 13 STATE: IL
 14 COUNTRY: USA
 15 ZIP: 60064-3500
 16 COMPUTER READABLE FORM:
 17 MEDIUM TYPE: Diskette
 18 COMPUTER: IBM Compatible
 19 OPERATING SYSTEM: DOS
 20 SOFTWARE: FastSeq for Windows Version 2.0
 21 CURRENT APPLICATION DATA:
 22 APPLICATION NUMBER: US/08/944,483
 23 FILING DATE:
 24 CLASSIFICATION: 424
 25 PRIOR APPLICATION DATA:
 26 APPLICATION NUMBER:
 27 FILING DATE:
 28 ATTORNEY/AGENT INFORMATION:
 29 NAME: Becker, Cheryl L.
 30 REGISTRATION NUMBER: 35,441
 31 REFERENCE/DOCKET NUMBER: 6183.US.01
 32 TELECOMMUNICATION INFORMATION:
 33 TELEPHONE: 847/935-1729
 34 TELEFAX: 847/938-2623
 35 TELEX:
 36 INFORMATION FOR SEQ ID NO: 63:
 37 SEQUENCE CHARACTERISTICS:
 38 LENGTH: 248 amino acids
 39 TYPE: amino acid
 40 STRANDEDNESS: single
 41 TOPOLOGY: linear
 42 MOLECULE TYPE: NO. 6232456e
 43 US-08-944-483-63

Query Match	42.0%	Score 553.5	DB 4	Length 248
Best Local Similarity	43.2%	Pred. No. 8.4e-54		
Matches	105	Conservative	46	Mismatches 83
				Indels 9
				Gaps 6

QY	2	IVGSGALPGAMPQVSLHVQ---	NVHVCGSGSIINPEVLTAAHCE-KPLNNPHMTAF	57
		: :	: : :	
Db	1	IVGTLNSSMGEMPMQVSLQVKLTAA	RHLGSGSLIGHQVLTAAHCEFDGLPDQVNR--	1Y 58
		: :	: : :	
QY	58	AGILRGSFMYGAGY-QVEKVISHP	NYDSKTKNNIDIALMLKLOKPLTFNDLVKPYCLP	116
		: :	: : :	
Db	59	SGILNLSDITKDTPEPQKEITIIHQ	NKVSSEGNHDIALIKLOAPLNTAEFGKPCILPS	118
		: :	: : :	
QY	117	MMLOPEQLCMISGCAATEBEKKTSE	VLNAAKVLLLETRCNSRYVYDNLITPAMTICAGEL	176
		: :	: : :	
Db	119	DSTIYTCWVYTWGMSFSEKGEIQ	NIILQKVINPLVTNECQKRY-QDYKITQRYMACAGY	177
		: :	: : :	
QY	177	OGNVDSCGGDSGGLPVYTSKNNIT	MLIGDTSMGSCGAAKAYRPVYGNVAVFDDMITRQMR-	235
		: :	: : :	
Db	178	EGGKDACGKGDGSGPLVCKHNGMM	RLVGLTISMGECARREOPGVYTKVAEYMDMILEKTQS	237
		: :	: : :	
QY	236	ADG 238		
Db	238	SDG 240		

RESULT	4
US-08-681-151-3	
: Sequence 3, Application US/08681151	

Db 1 IVGSDSREGAMPVVALYFDDOVCGASLVSRDMLVSAHCVGRNMESKMAVGLH 60
QY 62 ROSFMFYAGAYOVE-----KVISHPNYDSKTKNDIALMKLOKPLTFNDLVKPVCLPNPG 116
Db 61 MASNL-----TSPOIETRLDOIYINPHYKRRKNNDIAMHLEMKVNYDYIOPICLPEEN 117
QY 117 MMLQPEQLCWMISGWCATEERKKTSEVLNAAKVLLIETORCNSRYVYDNLITPAMICAGFL 176
Db 118 QVFPGRICSTINGCALIYGSTADVLQEDADVPLTSENKQCOQMPDEVN-ITENMYCAGYE 176
QY 177 QGNVDSGCGDSGGLVTSKNNIMMLIGDTSWGSCKAKATRPVYGNVAFWTMI 230
Db 177 AGGVDSGCGDSGGLMCGENRMMLAGVTSFGYQCALPNRPGVYARVPRTEMI 230

RESULT 11
US-09-008-271A-6
Sequence 6, Application US/09008271A
Patent No. 6203979
GENERAL INFORMATION:
APPLICANT: Bandman, Olga
Hillman, Jennifer L.
Yue, Henry
Guegler, Karl J.
Corley, Neil C.
Tang, Tom Y.
Shah, Purvi
TITLE OF INVENTION: HUMAN PROTEASE MOLECULES
NUMBER OF SEQUENCES: 24
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Dr.
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/008,271A
FILING DATE: 16-Jan-1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: <Unknown>
FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Mohan-Peterson, Sheela
REGISTRATION NUMBER: 41,201
REFERENCE/DOCKET NUMBER: PF-0458 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-855-0555
TELEFAX: 650-845-4166
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 435 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: COLNNOT13
CLONE: 1337018
SEQUENCE DESCRIPTION: SEQ ID NO: 6 :
US-09-008-271A-6

Query Match 39.5%; Score 521; DB 4; Length 435;
Best Local Similarity 44.6%; Pred. No. 8e-50;
Matches 108; Conservative 36; Mismatches 84; Indels 14; Gaps 7;
QY 1 RIVGESALPQAMPQVSLVQNVHVCGSLITPEWITAAHCVEKPLNNPMHTAFAGI 60
||||| :||||| :||||| :||||| :|||

Db 202 RVVGESEASVDSMPQVSVIQYDKQHYVCGSLDPHNVLTAAHCFRKH--TDVFNMKVRAGS 260
QY 61 LR-OSFMFYAGAYOVEKYI---SHPNYDSKTKNDIALMKLOKPLTFNDLVKPVCLPNPG 116
Db 261 DKLGSF-----PSLAVALKIIIFENPMY---PKNDIALMKLOKPLTFSGYVAPICLPFFD 313
QY 117 MMLQPEQLCWMISGWCATEER-KKTSEVLNAAKVLLIETORCNSRYVYDNLITPAMICAGF 175
Db 314 EELTPATPLMITIGWGTQKNGCKMSDILLQASVGYIDSTRCADAYAGEVTEKMKAGI 373
QY 176 LOGNVDSGCGDSGGLVTSKNNIMMLIGDTSWGSCKAKATRPVYGNVAFWTMI 235
Db 374 PEGGVDTGCGDSGGLMCGENRMMLAGVTSFGYQCALPNRPGVYARVPRTEMI 432
QY 236 AD 237
Db 433 AE 434

RESULT 12
US-08-508-448C-25
Sequence 25, Application US/08508448C
Patent No. 5804410
GENERAL INFORMATION:
APPLICANT: Kazuyoshi YAMAKA et al.
TITLE OF INVENTION: NUCLEIC ACID SEQUENCE ENCODING
NUMBER OF SEQUENCES: 25
CORRESPONDENCE ADDRESS:
ADDRESSEE: Wenderoth, Lind & Ponack
STREET: 805 Fifteenth Street, N.W., #700
CITY: Washington
STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20005
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.5 inch, 1.44 mb
COMPUTER: IBM Compatible
OPERATING SYSTEM: MS-DOS
SOFTWARE: Wordperfect 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/508,448C
FILING DATE: July 28, 1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Warren M. Cheek, Jr.
REGISTRATION NUMBER: 33,367
REFERENCE/DOCKET NUMBER:
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-371-8850
TELEFAX:
TEXT:
INFORMATION FOR SEQ ID NO: 25:
SEQUENCE CHARACTERISTICS:
LENGTH: 418 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-508-448C-25

Query Match 39.3%; Score 518.5; DB 1; Length 418;
Best Local Similarity 43.2%; Pred. No. 1.4e-49;
Matches 101; Conservative 34; Mismatches 94; Indels 5; Gaps 3;
QY 1 RIVGESALPQAMPQVSLVQNVHVCGSLITPEWITAAHCVEKPLNNPMHTAFAGI 60
||||| :||||| :||||| :||||| :|||
Db 186 RILGTEAFEGSWPQVSLRLNNAHHCGLINNMILTAAHCF-RSSNSNRDWTATSGI 244
QY 61 ROSFMFYAGAYOVEKVISHPNYDSKTKNDIALMKLOKPLTFNDLVKPVCLPNPGMLQ 120

Db 245 ---STFPLRRNRNLLHNHNSKATHEMDIALVRLNSVTFKDIHSVCLPATONIP 301
QY 121 PDLCHMSWGATTEEGKTSVYLNAAKVLLEFQRCNSRYVDNLITPAMICAGFLQGNV 180
Db 302 PSTAYVTGMAOEYAGHTVPELRQGVRIISNDVCNASHSYNGALLSGMLCAGVPOGGV 361
QY 181 DSCGDSGGPLV-TSKNNIMWLLIGDTSMGSCAKAYRPGVGNVVFETMIRQ 233
Db 362 DACGDSGGPLVQEDSRRLMFIVIGVSMGDCGLPDKPGVYTRVATYLDMIRQ 415

RESULT 13

US-08-508-448C-19
; Sequence 19, Application US/08508448C
; Patent No. 5804410
; GENERAL INFORMATION:
; APPLICANT: Kazuyoshi YAMAOKA et al.
; TITLE OF INVENTION: NUCLEIC ACID SEQUENCE ENCODING
; TITLE OF INVENTION: TRYPSIN-LIKE ENZYME AND PROCESS FOR PRODUCING THE ENZYME
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Wenderoth, Lind & Ponack
; STREET: 805 Fifteenth Street, N.W., #700
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 inch, 1.44 mb
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: Wordperfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/508.448C
; FILING DATE: July 28, 1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Warren M. Cheek, Jr.
; REGISTRATION NUMBER: 33,367
; REFERENCE/DOCKET NUMBER:
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-371-8850
; TELEFAX:
; TELEX:
; INFORMATION FOR SEQ ID NO: 19:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 232 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-508-448C-19

Query Match 38.9%; Score 513.5; DB 1; Length 232;

Best Local Similarity 42.9%; Pred. No. 2.2e-49;

Matches 100; Conservative 34; Mismatches 94; Indels 5; Gaps 3;

QY 2 YGGESALGAMPWOYSLHVNQVHVGSGSITPPEWIVTAACHCEKLPNNPMTAFAGLI 61
Db 1 IIGGTEAEGSNPMOYSLNNNAHCGSLNNMMLTAAHCF-KSNSNPRDIATSGI- 58
QY 62 ROSFMFYGAGYQVEKVISHPNDYDSTKNDIALMKLQKPLTFENDLVKPYCLPDPGMLAP 121
Db 59 --STFPLRRNRNLLHNHNSKATHEMDIALVRLNSVTFKDIHSVCLPATONIP 116
QY 122 EOLCHMSWGATTEEGKTSVYLNAAKVLLEFQRCNSRYVDNLITPAMICAGFLQGNV 181
Db 117 GSTAYVTGMAOEYAGHTVPELRQGVRIISNDVCNASHSYNGALLSGMLCAGVPOGGV 176

QY 182 SCGDSGGPLV-TSKNNIMWLLIGDTSMGSCAKAYRPGVGNVVFETMIRQ 233
Db 177 ACGDSGGPLVQEDSRRLMFIVIGVSMGDCGLPDKPGVYTRVATYLDMIRQ 229

RESULT 14

US-08-944-483-65
; Sequence 65, Application US/08944483
; Patent No. 6232456
; GENERAL INFORMATION:
; APPLICANT: COHEN, MAURICE
; APPLICANT: COLPITTS, TRACEY L.
; APPLICANT: FRIEDMAN, PAULA N.
; APPLICANT: GRANADOS, EDWARD N.
; APPLICANT: KLAAS, MICHAEL R.
; APPLICANT: RUSSELL, JOHN C.
; APPLICANT: STEWART, KENT D.
; APPLICANT: STROUPE, STEVEN D.
; TITLE OF INVENTION: NOVEL SERINE PROTEASE REAGENTS
; TITLE OF INVENTION: AND METHODS USEFUL FOR DETECTING AND TREATING DISEASES
; NUMBER OF SEQUENCES: 76
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Abbott Laboratories
; STREET: 100 Abbott Park Road
; CITY: Abbott Park
; STATE: IL
; COUNTRY: USA
; ZIP: 60064-3500
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/944.483
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Becker, Cheryl L.
; REGISTRATION NUMBER: 35,441
; REFERENCE/DOCKET NUMBER: 6183.US.01
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 847/935-1729
; TELEFAX: 847/938-2623
; TELEX:
; INFORMATION FOR SEQ ID NO: 65:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 235 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: No. 6232456e
US-08-944-483-65

Query Match 38.8%; Score 511.5; DB 4; Length 235;

Best Local Similarity 41.6%; Pred. No. 3.8e-49;

Matches 96; Conservative 46; Mismatches 86; Indels 3; Gaps 2;

QY 2 YGGESALGAMPWOYSLHVNQVHVGSGSITPPEWIVTAACHCEKLPNNPMTAFAGLI 61
Db 1 IYGSNAKEGAMPWVVGGLYGGRLCGASLVSSDWLVSANHCYVGNRLSPSKWTALIGH 60
QY 62 ROSFMF--YGAGYQVEKVISHPNDYDSTKNDIALMKLQKPLTFENDLVKPYCLPDPGML 119
Db 61 MKNLITSPQTVPLIDEIVINPYNRRKNDIATAMHLEKVVYTYIOPTCLPEENQYF 120
QY 120 QPDLCHMSWGATTEEGKTSVYLNAAKVLLEFQRCNSRYVDNLITPAMICAGFLQGNV 179

Db 121 PPRNCSTIAGWGTVVYQCTTANITLQADVPILISNRCQQQMPENV-ITENMICAGYEEGG 179
Qy 180 VDSGCGDSGGLVYTSKNNIMWLIGDTSWGSCKAKAYRPGVYGNVYFTDWI 230
Db 180 IDSCGDSGGLVYTSKNNIMWLIGDTSWGSCKAKAYRPGVYGNVYFTDWI 230

RESULT 15

US-09-000-846-2
: Sequence 2, Application us/09000846
: Patent No. 5981830
: GENERAL INFORMATION:
: APPLICANT: WU, QINGYU
: APPLICANT: SADDLER, JASPER
: TITLE OF INVENTION: KNOCKOUT MICE AND THEIR PROGENY WITH
: TITLE OF INVENTION: DISRUPTED SERINE PROTEASE GENES
: NUMBER OF SEQUENCES: 8
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: MILLEN, WHITE, ZELANO & BRANIGAN, P. C.
: STREET: 2200 CLARENDON BLVD, SUITE 1400
: CITY: ARLINGTON
: STATE: VA
: COUNTRY: US
: ZIP: 22201
: COMPUTER READABLE FORM:
: MEDIUM TYPE: floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/000.846
: FILING DATE: 30-DEC-1997
: CLASSIFICATION: 800
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/866,058
: FILING DATE: 30-MAY-1997
: ATTORNEY/AGENT INFORMATION:
: NAME: LEBOVITZ, RICHARD M.
: REGISTRATION NUMBER: 37,067
: REFERENCE/DOCKET NUMBER: BERLX 65P1
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 703-243-6333
: TELEFAX: 703-243-6410
: INFORMATION FOR SEQ ID NO: 2:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 416 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: US-09-000-846-2

Query Match 38.5%; Score 507.5; DB 2; Length 416;
Best Local Similarity 40.4%; Pred. No. 2.4e-48;
Matches 99; Conservative 42; Mismatches 93; Indels 11; Gaps 3;

Qy 1 RIVGESALPGAMPWOVSLHVNQNVHVCSSITPEVITAAHCVEKPLNPNHWTAFAGI 60
Db 161 RIVGGQDSGLRWPMQVSLRYDGTGLCGSLSGDWVLTAAHCFPERNRVLSRWRFAGA 220
Qy 61 LRQSFMEFGAGYQVEKYSHPNY-----DSKTKNDIALMKLQKPLTFNDLVKPCLPN 114
Db 221 VARTSP-HAVOLGQAVLYYHGGYLPFRDPTIDENSNDIALVHLSLPLTEYIQVCLPA 279
Qy 115 PGMLQPEQLCMISGMGATEEKGKTSVNLNAKAVLLIETORCNSRYVYDNLITPAMICAG 174
Db 280 AGQALVDBKCTVYGMGNTQYFGQAMVLOEARPIISNEVCNSPDFYGNQIKPMFCAG 339
Qy 175 FLAGNVDSGCGDSGGLV---TSKNNIMWLIGDTSWGSCKAKAYRPGVYGNVYFTDWI 230
Db 340 YPEGIDACGDSGGLVCFYCEDSISGTSRWRLCGIYSMCTGCALARKKPGVYTKVTDREMI 399
Qy 231 YRQMR 235

Db 400 FKAIR 404

Search completed: September 26, 2001, 09:22:09
Job time: 64 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 26, 2001, 09:26:15 : Search time 164.81 Seconds
(without alignments)
320.367 Million cell updates/sec

Title: US-09-615-285-2_COPY_255_492

Perfect score: 1319
Sequence: 1 RIVGSGSALPGAMPQVSLH.....VCGNVWFYFDWIRQMRAD 238

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 2803329 seqs, 221847457 residues

Total number of hits satisfying chosen parameters: 2803329

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 100%

Maximum Match 100%

Listing first 45 summaries

Database :

Pending_Patents_AA_Main:*

1: /cgn2_6/ptodata/2/paa/PCRTUS.COMB.pep:*
2: /cgn2_6/ptodata/2/paa/US06.COMB.pep:*
3: /cgn2_6/ptodata/2/paa/US07.COMB.pep:*
4: /cgn2_6/ptodata/2/paa/US08.COMB.pep:*
5: /cgn2_6/ptodata/2/paa/US081.COMB.pep:*
6: /cgn2_6/ptodata/2/paa/US082.COMB.pep:*
7: /cgn2_6/ptodata/2/paa/US083.COMB.pep:*
8: /cgn2_6/ptodata/2/paa/US084.COMB.pep:*
9: /cgn2_6/ptodata/2/paa/US085.COMB.pep:*
10: /cgn2_6/ptodata/2/paa/US086.COMB.pep:*
11: /cgn2_6/ptodata/2/paa/US087.COMB.pep:*
12: /cgn2_6/ptodata/2/paa/US089.COMB.pep:*
13: /cgn2_6/ptodata/2/paa/US090.COMB.pep:*
14: /cgn2_6/ptodata/2/paa/US091.COMB.pep:*
15: /cgn2_6/ptodata/2/paa/US092.COMB.pep:*
16: /cgn2_6/ptodata/2/paa/US093.COMB.pep:*
17: /cgn2_6/ptodata/2/paa/US094.COMB.pep:*
18: /cgn2_6/ptodata/2/paa/US095.COMB.pep:*
19: /cgn2_6/ptodata/2/paa/US096.COMB.pep:*
20: /cgn2_6/ptodata/2/paa/US097.COMB.pep:*
21: /cgn2_6/ptodata/2/paa/US098.COMB.pep:*
22: /cgn2_6/ptodata/2/paa/US099.COMB.pep:*
23: /cgn2_6/ptodata/2/paa/US560.COMB.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	1319	100.0	393 21 US-09-759-143-934	Sequence 934, App
2	1319	100.0	393 21 US-09-780-669-934	Sequence 934, App
3	1319	100.0	393 22 US-09-822-827-934	Sequence 934, App
4	1319	100.0	393 22 US-09-852-911-934	Sequence 934, App
5	1319	100.0	491 17 US-09-323-597-2	Sequence 2, Appli
6	1319	100.0	492 1 PCT-US99-14622-2	Sequence 2, Appli
7	1319	100.0	492 13 US-08-943-570-21	Sequence 21, Appli
8	1319	100.0	492 17 US-08-323-597B-2	Sequence 2, Appli
9	1319	100.0	492 20 US-09-615-285-2	Sequence 2, Appli
10	1319	100.0	492 20 US-09-691-840-2	Sequence 2, Appli

11	1319	100.0	492 21	US-09-759-143-932	Sequence 932, App
12	1319	100.0	492 21	US-09-780-669-932	Sequence 932, App
13	1319	100.0	492 22	US-09-822-827-932	Sequence 932, App
14	1319	100.0	492 22	US-09-852-911-932	Sequence 932, App
15	1316	99.8	492 18	US-09-408-628A-6	Sequence 6, Appli
16	1314	99.6	384 23	US-60-239-841-23	Sequence 23, Appli
17	1310	99.3	283 18	US-09-478-957-1	Sequence 1, Appli
18	1303	98.8	250 1	PCT-US00-06111-8	Sequence 8, Appli
19	1303	98.8	266 23	US-60-207-315-390	Sequence 1859, App
20	1303	98.8	266 23	US-60-230-435-1859	Sequence 15, Appli
21	1303	98.8	492 1	PCT-US99-19655-15	Sequence 5, Appli
22	1303	98.8	492 15	US-09-164-159-5	Sequence 4, Appli
23	1303	98.8	492 17	US-09-323-597-4	Sequence 4, Appli
24	1303	98.8	492 17	US-09-323-597B-4	Sequence 5, Appli
25	1303	98.8	492 18	US-09-408-628-5	Sequence 4, Appli
26	1303	98.8	492 20	US-09-615-285-4	Sequence 895, App
27	1303	98.8	492 20	US-09-679-426-895	Sequence 895, App
28	1303	98.8	492 20	US-09-685-166-895	Sequence 895, App
29	1303	98.8	492 21	US-09-759-143-895	Sequence 70, Appli
30	1303	98.8	492 21	US-09-776-191-70	Sequence 85, App
31	1303	98.8	492 21	US-09-780-669-895	Sequence 85, App
32	1303	98.8	492 22	US-09-822-827-895	Sequence 85, App
33	1303	98.8	492 22	US-09-852-911-895	Sequence 85, App
34	1303	98.8	492 22	US-09-879-792-14	Sequence 14, Appli
35	1303	98.8	605 23	US-60-212-659-375	Sequence 375, App
36	1303	98.8	672 23	US-60-212-659-373	Sequence 373, App
37	1090	82.6	227 23	US-60-230-435-1829	Sequence 1829, App
38	866	65.7	159 1	PCT-US00-05612-24	Sequence 24, Appli
39	866	65.7	159 19	US-09-518-046-24	Sequence 24, Appli
40	866	65.7	159 20	US-09-650-371-24	Sequence 24, Appli
41	655	49.7	453 17	US-09-380-138-69	Sequence 69, Appli
42	655	49.7	453 20	US-09-633-300-2	Sequence 2, Appli
43	655	49.7	453 21	US-09-776-191-12	Sequence 12, Appli
44	655	49.7	454 1	PCT-US00-05612-2	Sequence 2, Appli
45	655	49.7	454 19	US-09-518-046-2	Sequence 2, Appli

ALIGNMENTS

RESULT 1
US-09-759-143-934
Sequence 934, Application US/09759143
GENERAL INFORMATION:
APPLICANT: Xu, Jiangchun
APPLICANT: Dillon, Davin C.
APPLICANT: Mitcham, Jennifer L.
APPLICANT: Harlocker, Susan L.
APPLICANT: Jiang, Yuqi
APPLICANT: Henderson, Robert A.
APPLICANT: Kados, Michael D.
APPLICANT: Fanger, Gary R.
APPLICANT: Retter, Marc W.
APPLICANT: Stolk, John A.
APPLICANT: Day, Craig H.
APPLICANT: Vedvick, Thomas S.
APPLICANT: Carter, Darrick
APPLICANT: Li, Samuel
APPLICANT: Wang, Aijun
APPLICANT: Skeiky, Yasir A.W.
TITLE OF INVENTION: Heppler, William
TITLE OF INVENTION: COMPOSITIONS OF PROSTATE CANCER
FILE REFERENCE: 210121.427C23
CURRENT APPLICATION NUMBER: US/09/759,143
CURRENT FILING DATE: 2001-01-12
NUMBER OF SEQ ID NOS: 934
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 934
LENGTH: 393
TYPE: PRT
ORGANISM: Homo sapiens
US-09-759-143-934

Query Match 100.0%; Score 1319; DB 21; Length 393;
Best Local Similarity 100.0%; Pred. No. 5e-131;
Matches 238; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 RIVGESALPGAMPQVSLHVQNVHVCSSITTPEMITAAHCVEKPLNNPMHTAFAGI 60
DB 156 RIVGESALPGAMPQVSLHVQNVHVCSSITTPEMITAAHCVEKPLNNPMHTAFAGI 215
OY 61 LROSFMEYAGAYOEKVISHPNDYDSTKNNNDIALMKLQPLTFNDLVKPVCLPMPGMMLO 120
DB 216 LROSFMEYAGAYOEKVISHPNDYDSTKNNNDIALMKLQPLTFNDLVKPVCLPMPGMMLO 275
OY 121 PEOLCWSMGATEEKGKTSSEVLNAKVLLETORCNSRYVDNLITPAMICAGFLOGNV 180
DB 276 PEOLCWSMGATEEKGKTSSEVLNAKVLLETORCNSRYVDNLITPAMICAGFLOGNV 335
OY 181 DSCGDSGGLPVTSKNNIMWLIGDTSMGSCAKARPGVGNVVFDTWIROMRADG 238
DB 336 DSCGDSGGLPVTSKNNIMWLIGDTSMGSCAKARPGVGNVVFDTWIROMRADG 393

RESULT 2

US-09-780-669-934
Sequence 934, Application US/09780669
GENERAL INFORMATION:
APPLICANT: Xu, Jiangchun
APPLICANT: Dillon, Davin C.
APPLICANT: Mitcham, Jennifer L.
APPLICANT: Harlocker, Susan L.
APPLICANT: Jlang, Yuqi
APPLICANT: Henderson, Robert A.
APPLICANT: Kalos, Michael D.
APPLICANT: Fanger, Gary R.
APPLICANT: Retter, Marc W.
APPLICANT: Stolck, John A.
APPLICANT: Day, Craig H.
APPLICANT: Vedvick, Thomas S.
APPLICANT: Carter, Darlick
APPLICANT: Li, Samuel
APPLICANT: Wang, Aijun
APPLICANT: Skeiky, Yasir A.W.
APPLICANT: Hepler, William
APPLICANT: Hural, John
APPLICANT: McNeill, Patricia D.
APPLICANT: Houghton, Raymond L.
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
FILE REFERENCE: 210121.427C24
CURRENT APPLICATION NUMBER: US/09/780.669
CURRENT FILING DATE: 2001-02-09
NUMBER OF SEQ ID NOS: 943
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 934
LENGTH: 393
TYPE: PRT
ORGANISM: Homo sapiens
US-09-780-669-934

Query Match 100.0%; Score 1319; DB 21; Length 393;
Best Local Similarity 100.0%; Pred. No. 5e-131;
Matches 238; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 RIVGESALPGAMPQVSLHVQNVHVCSSITTPEMITAAHCVEKPLNNPMHTAFAGI 60
DB 156 RIVGESALPGAMPQVSLHVQNVHVCSSITTPEMITAAHCVEKPLNNPMHTAFAGI 215
OY 61 LROSFMEYAGAYOEKVISHPNDYDSTKNNNDIALMKLQPLTFNDLVKPVCLPMPGMMLO 120
DB 216 LROSFMEYAGAYOEKVISHPNDYDSTKNNNDIALMKLQPLTFNDLVKPVCLPMPGMMLO 275

OY 121 PEOLCWSMGATEEKGKTSSEVLNAKVLLETORCNSRYVDNLITPAMICAGFLOGNV 180
DB 276 PEOLCWSMGATEEKGKTSSEVLNAKVLLETORCNSRYVDNLITPAMICAGFLOGNV 335
OY 181 DSCGDSGGLPVTSKNNIMWLIGDTSMGSCAKARPGVGNVVFDTWIROMRADG 238
DB 336 DSCGDSGGLPVTSKNNIMWLIGDTSMGSCAKARPGVGNVVFDTWIROMRADG 393

RESULT 3

US-09-822-827-934
Sequence 934, Application US/09822827
GENERAL INFORMATION:
APPLICANT: Xu, Jiangchun
APPLICANT: Dillon, Davin C.
APPLICANT: Mitcham, Jennifer L.
APPLICANT: Harlocker, Susan L.
APPLICANT: Jlang, Yuqi
APPLICANT: Henderson, Robert A.
APPLICANT: Kalos, Michael D.
APPLICANT: Fanger, Gary R.
APPLICANT: Retter, Marc W.
APPLICANT: Stolck, John A.
APPLICANT: Day, Craig H.
APPLICANT: Vedvick, Thomas S.
APPLICANT: Carter, Darlick
APPLICANT: Li, Samuel
APPLICANT: Wang, Aijun
APPLICANT: Skeiky, Yasir A.W.
APPLICANT: Hepler, William
APPLICANT: Hural, John
APPLICANT: McNeill, Patricia D.
APPLICANT: Houghton, Raymond L.

Query Match 100.0%; Score 1319; DB 22; Length 393;
Best Local Similarity 100.0%; Pred. No. 5e-131;
Matches 238; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 RIVGESALPGAMPQVSLHVQNVHVCSSITTPEMITAAHCVEKPLNNPMHTAFAGI 60
DB 156 RIVGESALPGAMPQVSLHVQNVHVCSSITTPEMITAAHCVEKPLNNPMHTAFAGI 215
OY 61 LROSFMEYAGAYOEKVISHPNDYDSTKNNNDIALMKLQPLTFNDLVKPVCLPMPGMMLO 120
DB 216 LROSFMEYAGAYOEKVISHPNDYDSTKNNNDIALMKLQPLTFNDLVKPVCLPMPGMMLO 275
OY 121 PEOLCWSMGATEEKGKTSSEVLNAKVLLETORCNSRYVDNLITPAMICAGFLOGNV 180
DB 276 PEOLCWSMGATEEKGKTSSEVLNAKVLLETORCNSRYVDNLITPAMICAGFLOGNV 335
OY 181 DSCGDSGGLPVTSKNNIMWLIGDTSMGSCAKARPGVGNVVFDTWIROMRADG 238
DB 336 DSCGDSGGLPVTSKNNIMWLIGDTSMGSCAKARPGVGNVVFDTWIROMRADG 393

RESULT 4

US-09-852-911-934
Sequence 934, Application US/09852911
GENERAL INFORMATION:
APPLICANT: Xu, Jiangchun
APPLICANT: Dillon, Davin C.
APPLICANT: Mitcham, Jennifer L.
APPLICANT: Harlocker, Susan L.
APPLICANT: Jlang, Yuqi
APPLICANT: Henderson, Robert A.
APPLICANT: Kalos, Michael D.
APPLICANT: Fanger, Gary R.
APPLICANT: Retter, Marc W.
APPLICANT: Stolck, John A.
APPLICANT: Day, Craig H.
APPLICANT: Vedvick, Thomas S.
APPLICANT: Carter, Darlick
APPLICANT: Li, Samuel
APPLICANT: Wang, Aijun
APPLICANT: Skeiky, Yasir A.W.
APPLICANT: Hepler, William
APPLICANT: Hural, John
APPLICANT: McNeill, Patricia D.
APPLICANT: Houghton, Raymond L.

```

; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; TITLE OF INVENTION: DIAGNOSIS OF PROSTATE CANCER
; FILE REFERENCE: 210121.427C25
; CURRENT APPLICATION NUMBER: US/09/852,911
; CURRENT FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 947
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 934
; LENGTH: 393
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-852-911-934

Query Match          100.0%; Score 1319; DB 22; Length 393;
Best Local Similarity 100.0%; Pred. No. 5e-131;
Matches 238; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RIVGGSALPGAMPWOVSLHVONVHVGGSITTPETWITVAHCEYERKPLNNPMTAFAGI 60
    |||
DB 156 RIVGGSALPGAMPWOVSLHVONVHVGGSITTPETWITVAHCEYERKPLNNPMTAFAGI 215
    |||
QY 61 LROSEMFYAGYOVEKVISHPNDTSKTKNDIALMKLQKPLTFNDLVKPYCLPFGMLQ 120
    |||
DB 216 LROSEMFYAGYOVEKVISHPNDTSKTKNDIALMKLQKPLTFNDLVKPYCLPFGMLQ 275
    |||
QY 121 PEOLCWTISGCGATEEKGKTSSEVLNAKVLLEIQRCSRYVNDLITPAMICAGFLOGNV 180
    |||
DB 276 PEOLCWTISGCGATEEKGKTSSEVLNAKVLLEIQRCSRYVNDLITPAMICAGFLOGNV 335
    |||
QY 181 DSCGDSGGLVTSKNNIMWLIGDTSWGSCKAKAYRPGVGNVFTDWIYRQMRADG 238
    |||
DB 336 DSCGDSGGLVTSKNNIMWLIGDTSWGSCKAKAYRPGVGNVFTDWIYRQMRADG 393
    |||

RESULT 5
US-09-323-597-2
; Sequence 2, Application US/09323597
; GENERAL INFORMATION:
; APPLICANT: Afari, Daniel E
; APPLICANT: Hubert, Rene S
; APPLICANT: Leong, Kahen
; APPLICANT: Raitano, Arthur B.
; APPLICANT: Saifran, Douglas C.
; TITLE OF INVENTION: NOVEL TUMOR ANTIGEN USEFUL IN DIAGNOSIS AND THERAPY OF
; TITLE OF INVENTION: PROSTATE AND COLON CANCER
; FILE REFERENCE: 1703-007.US1
; CURRENT APPLICATION NUMBER: US/09/323,597
; CURRENT FILING DATE: 1999-06-01
; EARLIER APPLICATION NUMBER: 60/087,598
; EARLIER FILING DATE: 1998-06-01
; EARLIER APPLICATION NUMBER: 60/091,474
; EARLIER FILING DATE: 1998-06-29
; EARLIER APPLICATION NUMBER: 60/129,521
; EARLIER FILING DATE: 1999-04-14
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 491
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-323-597-2

Query Match          100.0%; Score 1319; DB 17; Length 491;
Best Local Similarity 100.0%; Pred. No. 6.6e-131;
Matches 238; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RIVGGSALPGAMPWOVSLHVONVHVGGSITTPETWITVAHCEYERKPLNNPMTAFAGI 60
    |||
DB 254 RIVGGSALPGAMPWOVSLHVONVHVGGSITTPETWITVAHCEYERKPLNNPMTAFAGI 313
    |||
QY 61 LROSEMFYAGYOVEKVISHPNDTSKTKNDIALMKLQKPLTFNDLVKPYCLPFGMLQ 120
    |||
```

```

    |||
DB 314 LROSEMFYAGYOVEKVISHPNDTSKTKNDIALMKLQKPLTFNDLVKPYCLPFGMLQ 373
    |||
QY 121 PEOLCWTISGCGATEEKGKTSSEVLNAKVLLEIQRCSRYVNDLITPAMICAGFLOGNV 180
    |||
DB 374 PEOLCWTISGCGATEEKGKTSSEVLNAKVLLEIQRCSRYVNDLITPAMICAGFLOGNV 433
    |||
QY 181 DSCGDSGGLVTSKNNIMWLIGDTSWGSCKAKAYRPGVGNVFTDWIYRQMRADG 238
    |||
DB 434 DSCGDSGGLVTSKNNIMWLIGDTSWGSCKAKAYRPGVGNVFTDWIYRQMRADG 491
    |||

RESULT 6
PCT-US99-14622-2
; Sequence 2, Application PCT/US9914622
; GENERAL INFORMATION:
; APPLICANT: Wong, Alexander K.C.
; APPLICANT: Tavligian, Sean V.
; APPLICANT: Teng, David H.-F.
; APPLICANT: Myriad Genetics, Inc.
; TITLE OF INVENTION: TMRSS2 is a Tumor Suppressor
; FILE REFERENCE: 2318-202
; CURRENT APPLICATION NUMBER: PCT/US99/14622
; CURRENT FILING DATE: 1999-07-01
; EARLIER APPLICATION NUMBER: US 60/091,044
; EARLIER FILING DATE: 1998-06-29
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 492
; TYPE: PRT
; ORGANISM: Homo sapiens
PCT-US99-14622-2

Query Match          100.0%; Score 1319; DB 1; Length 492;
Best Local Similarity 100.0%; Pred. No. 6.6e-131;
Matches 238; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RIVGGSALPGAMPWOVSLHVONVHVGGSITTPETWITVAHCEYERKPLNNPMTAFAGI 60
    |||
DB 255 RIVGGSALPGAMPWOVSLHVONVHVGGSITTPETWITVAHCEYERKPLNNPMTAFAGI 314
    |||
QY 61 LROSEMFYAGYOVEKVISHPNDTSKTKNDIALMKLQKPLTFNDLVKPYCLPFGMLQ 120
    |||
DB 315 LROSEMFYAGYOVEKVISHPNDTSKTKNDIALMKLQKPLTFNDLVKPYCLPFGMLQ 374
    |||
QY 121 PEOLCWTISGCGATEEKGKTSSEVLNAKVLLEIQRCSRYVNDLITPAMICAGFLOGNV 180
    |||
DB 375 PEOLCWTISGCGATEEKGKTSSEVLNAKVLLEIQRCSRYVNDLITPAMICAGFLOGNV 434
    |||
QY 181 DSCGDSGGLVTSKNNIMWLIGDTSWGSCKAKAYRPGVGNVFTDWIYRQMRADG 238
    |||
DB 435 DSCGDSGGLVTSKNNIMWLIGDTSWGSCKAKAYRPGVGNVFTDWIYRQMRADG 492
    |||

RESULT 7
US-08-943-570-21
; Sequence 21, Application US/08943570
; GENERAL INFORMATION:
; APPLICANT: COHEN, MAURICE
; APPLICANT: COLPITTS, TRACEY L.
; APPLICANT: FRIEDMAN, PAULA N.
; APPLICANT: GRANADOS, EDWARD N.
; APPLICANT: KLASSE, MICHAEL R.
; APPLICANT: RUSSELL, JOHN C.
; APPLICANT: STEWART, KENT D.
; APPLICANT: STROUPE, STEVEN D.
; TITLE OF INVENTION: NOVEL SPRING PROTEASE REAGENTS
; TITLE OF INVENTION: AND METHODS USEFUL FOR DETECTING AND TREATING DISEASES
; NUMBER OF SEQUENCES: 74
; CORRESPONDENCE ADDRESS:
```

```

ADDRESS:  Abbott Laboratories
STREET:   100 Abbott Park Road
CITY:     Abbott Park
STATE:    IL
COUNTRY:  USA
ZIP:      60064-3500
COMPUTER READABLE FORM:
MEDIUM TYPE:  Diskette
COMPUTER:    IBM Compatible
OPERATING SYSTEM:  DOS
SOFTWARE:    FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER:  US/08/943.570
FILING DATE:
CLASSIFICATION:  435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME:  Becker, Cheryl L.
REGISTRATION NUMBER:  35,441
REFERENCE/DOCKET NUMBER:  6182, US.01
TELECOMMUNICATION INFORMATION:
TELEPHONE:  847/935-1729
TELEFAX:    847/938-2623
TELEX:
INFORMATION FOR SEQ ID NO:  21:
SEQUENCE CHARACTERISTICS:
LENGTH:  492 amino acids
TYPE:    amino acid
STRANDEDNESS:  single
TOPOLOGY:  linear
MOLECULE TYPE:  None
US-08-943-570-21

Query Match          100.0%  Score 1319:  DB 13:  Length 492;
Best Local Similarity 100.0%:  Pred. No. 6.6e-131;
Matches 238:  Conservative  0:  Mismatches  0:  Indels  0:  Gaps  0:

QY  1  RIVGGSALPGAMPQVSLVHQNVAHVGSGSIITPEMIVTAHCVKRLPNPWHMTAFAGI  60
DB  255  RIVGGSALPGAMPQVSLVHQNVAHVGSGSIITPEMIVTAHCVKRLPNPWHMTAFAGI  314

QY  61  LROSMFPGAGGVQVKVYSHPRYDSKTKNNDIALLKLLKRLPFNDLVKRVCLPNFGMMIO  120
DB  315  LROSMFPGAGGVQVKVYSHPRYDSKTKNNDIALLKLLKRLPFNDLVKRVCLPNFGMMIO  374

QY  121  PEOLCWMISGNGATEKEGKTSEVLANAKVLLIETORCNSRYVYDNLITPAMICAGFLQGNV  180
DB  375  PEOLCWMISGNGATEKEGKTSEVLANAKVLLIETORCNSRYVYDNLITPAMICAGFLQGNV  434

QY  181  DSCGGDSCGPLVTSKNNITWMLIGDTSKSGSCAKATRGVYGVNMFDTMTYRQMRADG  238
DB  435  DSCGGDSCGPLVTSKNNITWMLIGDTSKSGSCAKATRGVYGVNMFDTMTYRQMRADG  492

RESULT      8
US-09-323-597B-2
Sequence 2, Application US/09323597B
GENERAL INFORMATION:
APPLICANT:  Daniel E. Afar
APPLICANT:  Rene S. Hubert
APPLICANT:  Kahan Leong
APPLICANT:  Arthur B. Raitano
APPLICANT:  Douglas C. Saffran
TITLE OF INVENTION:  NOVEL TUMOR ANTIGEN USEFUL IN DIAGNOSIS
FILE REFERENCE:  129, 80S01
CURRENT APPLICATION NUMBER:  US/09/323,597B
CURRENT FILING DATE:  1999-06-01
PRIOR APPLICATION NUMBER:  60/087,598
PRIOR FILING DATE:  1998-06-01

```

```

; PRIOR APPLICATION NUMBER: 60/091,474
; PRIOR FILING DATE: 1998-06-29
; PRIOR APPLICATION NUMBER: 60/129,521
; PRIOR FILING DATE: 1999-04-14
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO: 2
; LENGTH: 492
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-323-597B-2

Query Match      100.0%; Score 1319; DB 17; Length 492;
Best Local Similarity 100.0%; Pred. No. 6,6e-131;
Matches 238; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 RIVGESALPGAMPQVSLHVONVHVCSSITTPREIVTAACHVEKRLNNPHMTAFAGI 60
        |||||||
DB      255 RIVGESALPGAMPQVSLHVONVHVCSSITTPREIVTAACHVEKRLNNPHMTAFAGI 314
        |||||||

QY      61 LRQSMFYGAGVQVEKVLISHPNYDSKTKNDIALAKLQRLPFENDLVKRVCLPNEGMMIQ 120
        |||||||
DB      315 LRQSMFYGAGVQVEKVLISHPNYDSKTKNDIALAKLQRLPFENDLVKRVCLPNEGMMIQ 374
        |||||||

QY      121 PRQLCWISGWTAEKTKTSEVLAARVLLIEFQRCNSRYVDNLITPAMICAGFLQGNV 180
        |||||||
DB      375 PRQLCWISGWTAEKTKTSEVLAARVLLIEFQRCNSRYVDNLITPAMICAGFLQGNV 434
        |||||||

QY      181 DSCQDSSGGLVTSKNNIMWLLIGDPSWGSgcAKAYRPGVYGVNVTETDIYROMRADG 238
        |||||||
DB      435 DSCQDSSGGLVTSKNNIMWLLIGDPSWGSgcAKAIRPGVYGVNVTETDIYROMRADG 492
        |||||||

RESULT          9
US-09-615-285-2
; Sequence 2, Application US/09615285
; GENERAL INFORMATION:
; APPLICANT: Daniel E.H. Afar
; APPLICANT: Rene S. Hubert
; APPLICANT: Kanun Leong
; APPLICANT: Arthur B. Raitano
; APPLICANT: Douglas C. Saffran
; APPLICANT: Stephen C. Mitchell
; APPLICANT: Aya Jakobovits
; APPLICANT: Mary Paris
; APPLICANT: Igor Vivanco
; TITLE OF INVENTION: NOVEL TUMOR ANTIGEN USEFUL IN DIAGNOSIS
; TITLE OF INVENTION: AND THERAPY OF PROSTATE AND COLON CANCER
; FILE REFERENCE: 129.90USII
; CURRENT APPLICATION NUMBER: US/09/615,285
; PRIOR APPLICATION NUMBER: 09/323,597
; PRIOR FILING DATE: 2000-07-12
; PRIOR FILING DATE: 1999-06-01
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO: 2
; LENGTH: 492
; TYPE: PRT
; ORGANISM: Homo Sapiens
US-09-615-285-2

Query Match      100.0%; Score 1319; DB 20; Length 492;
Best Local Similarity 100.0%; Pred. No. 6,6e-131;
Matches 238; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 RIVGESALPGAMPQVSLHVONVHVCSSITTPREIVTAACHVEKRLNNPHMTAFAGI 60
        |||||||
DB      255 RIVGESALPGAMPQVSLHVONVHVCSSITTPREIVTAACHVEKRLNNPHMTAFAGI 314
        |||||||

QY      61 LRQSMFYGAGVQVEKVLISHPNYDSKTKNDIALAKLQRLPFENDLVKRVCLPNEGMMIQ 120
        |||||||
DB      315 LRQSMFYGAGVQVEKVLISHPNYDSKTKNDIALAKLQRLPFENDLVKRVCLPNEGMMIQ 120
        |||||||

```



```

: TYPE: PRT
: ORGANISM: Homo sapiens
US-09-780-669-932

```

```

Query Match      100.0%; Score 1319; DB 21; Length 492;
Best Local Similarity 100.0%; Pred. No. 6.6e-131;
Matches 238; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Oy 1 RIVGESALPGAMPQVSLHVONVHVCSSITTPETVITAACHVEKRLNNPWHMTAFAGI 60
    |||||||
Db 255 RIVGESALPGAMPQVSLHVONVHVCSSITTPETVITAACHVEKRLNNPWHMTAFAGI 314
    |||||||
Oy 61 LROSFMFYGAGYQVEKVISHPNYDSKTKNDIALMKLQKPLTFNDLVKPVCLPFGMMLO 120
    |||||||
Db 315 LROSFMFYGAGYQVEKVISHPNYDSKTKNDIALMKLQKPLTFNDLVKPVCLPFGMMLO 374
    |||||||
Oy 121 PEOLCWMISGWTGTEKGTSEVLNAKAVLLIETORCNSRYVDNLITPAMICAGFLQGNV 180
    |||||||
Db 375 PEOLCWMISGWTGTEKGTSEVLNAKAVLLIETORCNSRYVDNLITPAMICAGFLQGNV 434
    |||||||
Oy 181 DSCGDSGGLVTSKNNIMWLIGDTSWGSCKAKAYRPGYGVNVFTDWTYROMRADG 238
    |||||||
Db 435 DSCGDSGGLVTSKNNIMWLIGDTSWGSCKAKAYRPGYGVNVFTDWTYROMRADG 492
    |||||||

```

RESULT 13

```

US-09-822-827-932
: Sequence 932, Application US/09822827
: GENERAL INFORMATION:

```

```

: APPLICANT: Xu, Jlangchun
: TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
: FILE REFERENCE: 210121.534C1
: CURRENT APPLICATION NUMBER: US/09/822,827
: CURRENT FILING DATE: 2001-03-28
: NUMBER OF SEQ ID NOS: 982
: SOFTWARE: FastSeq for Windows Version 3.0
: SEQ ID NO 932
: LENGTH: 492
: TYPE: PRT
: ORGANISM: Homo sapiens
US-09-822-827-932

```

```

Query Match      100.0%; Score 1319; DB 22; Length 492;
Best Local Similarity 100.0%; Pred. No. 6.6e-131;
Matches 238; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Oy 1 RIVGESALPGAMPQVSLHVONVHVCSSITTPETVITAACHVEKRLNNPWHMTAFAGI 60
    |||||||
Db 255 RIVGESALPGAMPQVSLHVONVHVCSSITTPETVITAACHVEKRLNNPWHMTAFAGI 314
    |||||||
Oy 61 LROSFMFYGAGYQVEKVISHPNYDSKTKNDIALMKLQKPLTFNDLVKPVCLPFGMMLO 120
    |||||||
Db 315 LROSFMFYGAGYQVEKVISHPNYDSKTKNDIALMKLQKPLTFNDLVKPVCLPFGMMLO 374
    |||||||
Oy 121 PEOLCWMISGWTGTEKGTSEVLNAKAVLLIETORCNSRYVDNLITPAMICAGFLQGNV 180
    |||||||
Db 375 PEOLCWMISGWTGTEKGTSEVLNAKAVLLIETORCNSRYVDNLITPAMICAGFLQGNV 434
    |||||||
Oy 181 DSCGDSGGLVTSKNNIMWLIGDTSWGSCKAKAYRPGYGVNVFTDWTYROMRADG 238
    |||||||
Db 435 DSCGDSGGLVTSKNNIMWLIGDTSWGSCKAKAYRPGYGVNVFTDWTYROMRADG 492
    |||||||

```

RESULT 14

```

US-09-852-911-932
: Sequence 932, Application US/09852911
: GENERAL INFORMATION:
: APPLICANT: Xu, Jlangchun
: APPLICANT: Dillon, David C.
: APPLICANT: Mitcham, Jennifer L.

```

```

: APPLICANT: Harlocker, Susan L.
: APPLICANT: Jiang, Yuqi
: APPLICANT: Henderson, Robert A.
: APPLICANT: Kalos, Michael D.
: APPLICANT: Fanger, Gary R.
: APPLICANT: Retter, Marc W.
: APPLICANT: Stolk, John A.
: APPLICANT: Day, Craig H.
: APPLICANT: Vedvick, Thomas S.
: APPLICANT: Carter, Darlick
: APPLICANT: Li, Samuel
: APPLICANT: Wang, Aijun
: APPLICANT: Skeiky, Yasir A.W.
: APPLICANT: Hepler, William
: APPLICANT: Hural, John
: APPLICANT: McNeill, Patricia D.
: APPLICANT: Houghton, Raymond L.
: TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
: FILE REFERENCE: 210121.427C25
: CURRENT APPLICATION NUMBER: US/09/852,911
: CURRENT FILING DATE: 2001-05-09
: NUMBER OF SEQ ID NOS: 947
: SOFTWARE: FastSeq for Windows Version 3.0
: SEQ ID NO 932
: LENGTH: 492
: TYPE: PRT
: ORGANISM: Homo sapiens
US-09-852-911-932

```

```

Query Match      100.0%; Score 1319; DB 22; Length 492;
Best Local Similarity 100.0%; Pred. No. 6.6e-131;
Matches 238; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Oy 1 RIVGESALPGAMPQVSLHVONVHVCSSITTPETVITAACHVEKRLNNPWHMTAFAGI 60
    |||||||
Db 255 RIVGESALPGAMPQVSLHVONVHVCSSITTPETVITAACHVEKRLNNPWHMTAFAGI 314
    |||||||
Oy 61 LROSFMFYGAGYQVEKVISHPNYDSKTKNDIALMKLQKPLTFNDLVKPVCLPFGMMLO 120
    |||||||
Db 315 LROSFMFYGAGYQVEKVISHPNYDSKTKNDIALMKLQKPLTFNDLVKPVCLPFGMMLO 374
    |||||||
Oy 121 PEOLCWMISGWTGTEKGTSEVLNAKAVLLIETORCNSRYVDNLITPAMICAGFLQGNV 180
    |||||||
Db 375 PEOLCWMISGWTGTEKGTSEVLNAKAVLLIETORCNSRYVDNLITPAMICAGFLQGNV 434
    |||||||
Oy 181 DSCGDSGGLVTSKNNIMWLIGDTSWGSCKAKAYRPGYGVNVFTDWTYROMRADG 238
    |||||||
Db 435 DSCGDSGGLVTSKNNIMWLIGDTSWGSCKAKAYRPGYGVNVFTDWTYROMRADG 492
    |||||||

```

RESULT 15

```

US-09-408-628A-6
: Sequence 6, Application US/09408628A
: GENERAL INFORMATION:
: APPLICANT: Macbeth, Kyle J.
: APPLICANT: Shyjan, Andrew W.
: TITLE OF INVENTION: NUCLEIC ACIDS AND POLYPEPTIDES USEFUL IN
: FILE REFERENCE: 0734-135001
: CURRENT APPLICATION NUMBER: US/09/408,628A
: CURRENT FILING DATE: 1998-09-30
: PRIOR APPLICATION NUMBER: US 09/164,159
: PRIOR FILING DATE: 1998-09-30
: PRIOR APPLICATION NUMBER: US 09/163,759
: PRIOR FILING DATE: 1998-09-30
: NUMBER OF SEQ ID NOS: 18
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 6
: LENGTH: 492
: TYPE: PRT
: ORGANISM: Homo sapiens

```

US-09-408-628A-6

Query Match 99.8%; Score 1316; DB 18; Length 492;
Best Local Similarity 99.6%; Pred. No. 1.4e-130;
Matches 237; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY	1	RIVGESALPGAMPQVSLHVQNVHVCSSITPEWIVTAHCEKPLNNPWHMTAFAGI	60
DB	255	RIVGESALPGAMPQVSLHVQNVHVCSSITPEWIVTAHCEKPLNNPWHMTAFAGI	314
QY	61	LROSEMFYAGYQVEKVISHPNYSKTKNNDIALLMKLQKPLTFNDLVKPYCLPMPGMLQ	120
DB	315	LROSEMFYAGYQVQKVISHPNYSKTKNNDIALLMKLQKPLTFNDLVKPYCLPMPGMLQ	374
QY	121	PROLCWISGNGATEEKGKTSVLNAKVLIIETQRCNSRYVDNLITPAMICAGFLQGNV	180
DB	375	PROLCWISGNGATEEKGKTSVLNAKVLIIETQRCNSRYVDNLITPAMICAGFLQGNV	434
QY	181	DSQGDGSGPLVTSKNNIMWLLIGDTSNGSGCAKAYRPGVYGNMVFPTDNIYROMRADG	238
DB	435	DSQGDGSGPLVTSKNNIMWLLIGDTSNGSGCAKAYRPGVYGNMVFPTDNIYROMRADG	492

Search completed: September 26, 2001, 09:26:16
Job time: 311 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 26, 2001, 09:23:22 : Search time 27.32 Seconds
(without alignments)
370.445 Million cell updates/sec

Title: US-09-615-285-2_COPY_255_492
Perfect score: 1319
Sequence: 1 RIVGESALPGAMPQVSLH.....VYGNVVFETDVIYRQMRADG 238

Scoring table: BIOSUM62
Gapop 10.0, Gapext 0.5

Searched: 277479 seqs, 42523304 residues

Total number of hits satisfying chosen parameters: 277479

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Pending_Patents_AA_New.*
1: /cgn2_6/ptodata/1/paa/PCF_NEW_COMB.pep.*
2: /cgn2_6/ptodata/1/paa/US06_NEW_COMB.pep.*
3: /cgn2_6/ptodata/1/paa/US07_NEW_COMB.pep.*
4: /cgn2_6/ptodata/1/paa/US08_NEW_COMB.pep.*
5: /cgn2_6/ptodata/1/paa/US09_NEW_COMB.pep.*
6: /cgn2_6/ptodata/1/paa/US60_NEW_COMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1319	100.0	260	US-09-760-475-2355	Sequence 2355, Ap
2	1319	100.0	393	US-09-895-793-934	Sequence 934, App
3	1319	100.0	393	US-09-895-814-934	Sequence 934, App
4	1319	100.0	492	US-09-895-793-932	Sequence 932, App
5	1319	100.0	492	US-09-895-814-932	Sequence 932, App
6	1303	98.8	492	US-09-895-793-895	Sequence 895, App
7	1303	98.8	492	US-09-895-814-895	Sequence 895, App
8	1263	95.8	229	US-09-715-427-53	Sequence 53, App
9	894	67.8	163	US-09-760-475-3414	Sequence 3414, Ap
10	645.5	48.9	248	US-09-925-301-1017	Sequence 1017, Ap
11	645.5	48.9	394	PCF-US01-18569-3126	Sequence 3126, Ap
12	561	42.5	421	US-09-715-427-15	Sequence 15, App
13	529.5	40.1	343	US-09-875-195-922	Sequence 922, App
14	523.5	39.7	255	US-09-410-362B-6	Sequence 6, App
15	521	39.5	435	PCF-US01-18568-2	Sequence 2, App
16	521	39.5	435	PCF-US01-18568A-2	Sequence 2, App
17	511.5	38.8	235	US-09-410-362B-5	Sequence 5, App
18	509	38.6	521	US-09-764-898-216	Sequence 216, App
19	506	38.4	292	PCF-US01-18568-9	Sequence 9, App
20	506	38.4	292	PCF-US01-18568A-9	Sequence 9, App
21	502.5	38.1	812	US-09-457-018-81	Sequence 81, App
22	502	38.1	236	US-09-715-427-44	Sequence 44, App
23	502	38.1	855	US-09-410-362B-2	Sequence 2, App
24	497	37.7	241	US-09-410-362B-3	Sequence 3, App
25	477	36.2	932	PCF-US01-08631-51801	Sequence 51801, A
26	466	35.3	916	US-09-758-449-1401	Sequence 1401, Ap
27	465.5	35.3	293	US-09-764-898-213	Sequence 213, Ap

28	465.5	35.3	449	1	PCF-US01-08631-58151	Sequence 58151, A
29	458	34.7	413	5	US-09-715-427-11	Sequence 11, App
30	457.5	34.7	305	5	US-09-885-441-57	Sequence 57, App
31	457.5	34.7	334	5	US-09-885-441-56	Sequence 56, App
32	455.5	34.5	244	5	US-09-715-427-51	Sequence 51, App
33	455.5	34.5	314	5	US-09-885-441-14	Sequence 14, App
34	455.5	34.5	317	5	US-09-664-610B-257	Sequence 257, App
35	455.5	34.5	317	5	US-09-882-636-23	Sequence 23, App
36	455.5	34.5	317	5	US-09-664-610B-263	Sequence 263, App
37	455.5	34.5	325	5	US-09-764-898-278	Sequence 278, App
38	455.5	34.5	325	5	US-09-764-902-1211	Sequence 1211, App
39	455.5	34.5	325	5	US-09-908-711-114	Sequence 114, App
40	455.5	34.5	327	5	US-09-787-844-2	Sequence 2, App
41	455.5	34.5	351	5	US-09-764-898-207	Sequence 207, App
42	454.5	34.5	322	5	US-09-764-898-279	Sequence 279, App
43	454.5	34.5	322	5	US-09-764-902-1747	Sequence 1747, App
44	452.5	34.3	230	5	US-09-760-475-2331	Sequence 2331, App
45	451	34.2	260	5	US-09-885-441-58	Sequence 58, App

ALIGNMENTS

```
RESULT 1
US-09-760-475-2355
; Sequence 2355, Application US/09760475
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT249
; CURRENT APPLICATION NUMBER: US/09/760,475
; PRIORITY FILING DATE: 2001-01-16
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 4122
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2355
; LENGTH: 260
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-760-475-2355

Query Match      100.0%; Score 1319; DB 5; Length 260;
Best Local Similarity 100.0%; Pred. No. 1.3e+120;
Matches 238; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RIVGESALPGAMPQVSLHVNVAHVGSGSITPEMITVAHCVCKPLNNPWHMTAFAGI 60
    |||||||
DB 23 RIVGESALPGAMPQVSLHVNVAHVGSGSITPEMITVAHCVCKPLNNPWHMTAFAGI 82
    |||||||

QY 61 LRQSFHFYAGYVEKVISHPNDSTKKNNDIALMLQRPFTNDLVKPYCLPQGMILQ 120
    |||||||
DB 83 LRQSFHFYAGYVEKVISHPNDSTKKNNDIALMLQRPFTNDLVKPYCLPQGMILQ 142
    |||||||

QY 121 PEOLCMTSGGATEEGKTSSEVINAIVLLIETQCNRSRYVYNLTTPAMICAGFLQGNV 180
    |||||||
DB 143 PEOLCMTSGGATEEGKTSSEVINAIVLLIETQCNRSRYVYNLTTPAMICAGFLQGNV 202
    |||||||

QY 181 DSCQGSGLPGVLTYSKNNIMWLIGDTSWGSGLCAKARPGYGVNMFETDIYRQMRADG 238
    |||||||
DB 203 DSCQGSGLPGVLTYSKNNIMWLIGDTSWGSGLCAKARPGYGVNMFETDIYRQMRADG 260
    |||||||

RESULT 2
US-09-895-793-934
; Sequence 934, Application US/09895793
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillon, David C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Jiang, Yugu
; APPLICANT: Kalos, Michael D.
```

```

: APPLICANT: Retter, Marc W.
: APPLICANT: Stolck, John A.
: APPLICANT: Day, Craig H.
: APPLICANT: Vedvick, Thomas S.
: APPLICANT: Carter, Darlick
: APPLICANT: Li, Samuel X.
: APPLICANT: Wang, Aljun
: APPLICANT: Skeiky, Yasir A.W.
: APPLICANT: Hepler, William T.
: APPLICANT: Henderson, Robert A.
: APPLICANT: Hural, John
: APPLICANT: McNeill, Patricia D.
: APPLICANT: Houghton, Raymond L.
: APPLICANT: Vinals de Bassols, Carlota
: APPLICANT: Foy, Teresa
: APPLICANT: Fanger, Gary R.
: TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
: TITLE OF INVENTION: DIAGNOSIS OF PROSTATE CANCER
: FILE REFERENCE: 210121.534C2
: CURRENT APPLICATION NUMBER: US/09/895,793
: CURRENT FILING DATE: 2001-06-29
: NUMBER OF SEQ ID NOS: 982
: SOFTWARE: FastSeq for Windows Version 3.0
: SEQ ID NO 934
: LENGTH: 393
: TYPE: PRT
: ORGANISM: Homo sapiens
US-09-895-793-934
```

```

Query Match          100.0%; Score 1319; DB 5; Length 393;
Best Local Similarity 100.0%; Pred. No. 2.2e-120;
Matches 238; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```

Oy 1 RIVGESALPGAMPQVSLHVNQNVHVCSSITTPKIVTAACHVEKPLNNPWHMTAFAGI 60
    |||||||
Db 156 RIVGESALPGAMPQVSLHVNQNVHVCSSITTPKIVTAACHVEKPLNNPWHMTAFAGI 215
    |||||||
Oy 61 LRQSFMYGAGYQVEKVI SHPNYDSKTKNDIALMLKQPLTFNDLVKPVCLPNEGMLQ 120
    |||||||
Db 216 LRQSFMYGAGYQVEKVI SHPNYDSKTKNDIALMLKQPLTFNDLVKPVCLPNEGMLQ 275
    |||||||
Oy 121 PEQLCWISGWTGTEKGTSEVLNAKAVLLIETQRCNSRYVDNLITPAMICAGFLQGNV 180
    |||||||
Db 276 PEQLCWISGWTGTEKGTSEVLNAKAVLLIETQRCNSRYVDNLITPAMICAGFLQGNV 335
    |||||||
Oy 181 DSCQDSSGPLVTSKNNIMWLIGDTSMGSGCAKAYRPGVYGVMVFTDWTYROMRADG 238
    |||||||
Db 336 DSCQDSSGPLVTSKNNIMWLIGDTSMGSGCAKAYRPGVYGVMVFTDWTYROMRADG 393
    |||||||
```

```

RESULT 3
US-09-895-814-934
: Sequence 934, Application US/09895814
: GENERAL INFORMATION:
: APPLICANT: Xu, Jlangchun
: APPLICANT: Dillon, Davin C.
: APPLICANT: Mitcham, Jennifer L.
: APPLICANT: Harlocker, Susan L.
: APPLICANT: Jlang, Yuguu
: APPLICANT: Kalos, Michael D.
: APPLICANT: Retter, Marc W.
: APPLICANT: Stolck, John A.
: APPLICANT: Day, Craig H.
: APPLICANT: Vedvick, Thomas S.
: APPLICANT: Carter, Darlick
: APPLICANT: Li, Samuel X.
: APPLICANT: Wang, Aljun
: APPLICANT: Skeiky, Yasir A.W.
: APPLICANT: Hepler, William T.
: APPLICANT: Henderson, Robert A.
: APPLICANT: Hural, John
: APPLICANT: McNeill, Patricia D.
```

```

: APPLICANT: Houghton, Raymond L.
: APPLICANT: Vinals de Bassols, Carlota
: APPLICANT: Foy, Teresa
: APPLICANT: Fanger, Gary R.
: TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
: TITLE OF INVENTION: DIAGNOSIS OF PROSTATE CANCER
: FILE REFERENCE: 210121.427C26
: CURRENT APPLICATION NUMBER: US/09/895,814
: CURRENT FILING DATE: 2001-06-29
: NUMBER OF SEQ ID NOS: 990
: SOFTWARE: FastSeq for Windows Version 3.0
: SEQ ID NO 934
: LENGTH: 393
: TYPE: PRT
: ORGANISM: Homo sapiens
US-09-895-814-934
```

```

Query Match          100.0%; Score 1319; DB 5; Length 393;
Best Local Similarity 100.0%; Pred. No. 2.2e-120;
Matches 238; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```

Oy 1 RIVGESALPGAMPQVSLHVNQNVHVCSSITTPKIVTAACHVEKPLNNPWHMTAFAGI 60
    |||||||
Db 156 RIVGESALPGAMPQVSLHVNQNVHVCSSITTPKIVTAACHVEKPLNNPWHMTAFAGI 215
    |||||||
Oy 61 LRQSFMYGAGYQVEKVI SHPNYDSKTKNDIALMLKQPLTFNDLVKPVCLPNEGMLQ 120
    |||||||
Db 216 LRQSFMYGAGYQVEKVI SHPNYDSKTKNDIALMLKQPLTFNDLVKPVCLPNEGMLQ 275
    |||||||
Oy 121 PEQLCWISGWTGTEKGTSEVLNAKAVLLIETQRCNSRYVDNLITPAMICAGFLQGNV 180
    |||||||
Db 276 PEQLCWISGWTGTEKGTSEVLNAKAVLLIETQRCNSRYVDNLITPAMICAGFLQGNV 335
    |||||||
Oy 181 DSCQDSSGPLVTSKNNIMWLIGDTSMGSGCAKAYRPGVYGVMVFTDWTYROMRADG 238
    |||||||
Db 336 DSCQDSSGPLVTSKNNIMWLIGDTSMGSGCAKAYRPGVYGVMVFTDWTYROMRADG 393
    |||||||
```

```

RESULT 4
US-09-895-793-932
: Sequence 932, Application US/09895793
: GENERAL INFORMATION:
: APPLICANT: Xu, Jlangchun
: APPLICANT: Dillon, Davin C.
: APPLICANT: Mitcham, Jennifer L.
: APPLICANT: Harlocker, Susan L.
: APPLICANT: Jlang, Yuguu
: APPLICANT: Kalos, Michael D.
: APPLICANT: Retter, Marc W.
: APPLICANT: Stolck, John A.
: APPLICANT: Day, Craig H.
: APPLICANT: Vedvick, Thomas S.
: APPLICANT: Carter, Darlick
: APPLICANT: Li, Samuel X.
: APPLICANT: Wang, Aljun
: APPLICANT: Skeiky, Yasir A.W.
: APPLICANT: Hepler, William T.
: APPLICANT: Henderson, Robert A.
: APPLICANT: Hural, John
: APPLICANT: McNeill, Patricia D.
: APPLICANT: Houghton, Raymond L.
: APPLICANT: Vinals de Bassols, Carlota
: APPLICANT: Foy, Teresa
: APPLICANT: Fanger, Gary R.
: TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
: TITLE OF INVENTION: DIAGNOSIS OF PROSTATE CANCER
: FILE REFERENCE: 210121.534C2
: CURRENT APPLICATION NUMBER: US/09/895,793
: CURRENT FILING DATE: 2001-06-29
: NUMBER OF SEQ ID NOS: 982
: SOFTWARE: FastSeq for Windows Version 3.0
: SEQ ID NO 932
```

```

; LENGTH: 492
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-895-793-932

```

Query Match	100.0%;	Score 1319;	DB 5;	Length 492;
Best Local Similarity	100.0%;	Pred. NO. 2.9e-120;		
Matches 238; Conservative	0;	Mismatches	0;	Indels 0;
				Gaps 0;

QY	1	RIVGESSLPAMPAMQVSLHJQNVHGGGSLTPEWLTVAHCEYKLNPNHMTAFAGI	60
Db	255	RIVGESSLPAMPAMQVSLHJQNVHGGGSLTPEWLTVAHCEYKLNPNHMTAFAGI	314
QY	61	LROSFMEFYAGYQVEKYLSHPNDVDSKTRNDIALMLQKPLFENDLVKPVCLPBNPKMLD	120
Db	315	LROSFMEFYAGYQVEKYLSHPNDVDSKTRNDIALMLQKPLFENDLVKPVCLPBNPKMLD	374
QY	121	PEOLCWLISGKQATEEKGKTSFVLLAAVLLIETORCNSRIVYDNLITPAMICAGFLQGNV	180
Db	375	PEOLCWLISGKQATEEKGKTSFVLLAAVLLIETORCNSRIVYDNLITPAMICAGFLQGNV	434
QY	181	DSOGGDSGGLVYTSKNNINIMWLIIGTSMGSGCARKYRPGVGNVVFETDYLIRQMRADG	238
Db	435	DSOGGDSGGLVYTSKNNINIMWLIIGTSMGSGCARKYRPGVGNVVFETDYLIRQMRADG	492

RESULT 5
US-09-895-814-932

```

APPLICANT: Xu, JIangchun
APPLICANT: Dillon, Davin C.
APPLICANT: Mitcham, Jennifer L.
APPLICANT: Harlocker, Susan L.
APPLICANT: Jiang, Yugu
APPLICANT: Kalos, Michael D.
APPLICANT: Ketter, Marc W.
APPLICANT: Stolk, John A.
APPLICANT: Day, Craig H.
APPLICANT: Vedvick, Thomas S.
APPLICANT: Carter, Darlick
APPLICANT: Li, Samuel X.
APPLICANT: Wang, Aijun
APPLICANT: Skeiky, Vasil A.W.
APPLICANT: Hepler, William F.
APPLICANT: Henderson, Robert A.
APPLICANT: Hural, John
APPLICANT: McNeill, Patricia D.
APPLICANT: Houghton, Raymond L.
APPLICANT: Vinals de Bassols, Carlota
APPLICANT: Foy, Teresa
APPLICANT: Panger, Gary R.
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
FILE REFERENCE: 210121.427C26
CURRENT APPLICATION NUMBER: US/09/895,814
CURRENT FILING DATE: 2001-06-29
NUMBER OF SEQ ID NOS: 90
SOFTWARE: FASTSEQ for Windows Version 3.0
SEQ ID NO 932
LENGTH: 492
TYPE: PRT
ORGANISM: Homo sapiens
US-09-895-814-932

```

```

Query Match      100.0%; Score 1319; DB 5; Length 492;
Best Local Similarity 100.0%; Pred. No. 2.9e-120;
Matches 238; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 RIVGESALPGAMPQVSLHVNQNVCGSITTPENIYTAHCVKRLNNPHWTFACI 60
|||||

```

Db	235	RIVGESSLPEAMWQWYSLHJQVNHVGGSIITPEWIVTFAAHCEKRLNPNHHTATAGI	314
Qy	61	LROSFMEFYGAGYQVEKVISHPNDYDSKTKNNIDIALMLQKPLTFENDLVKPYCLPBPNGMLQ	120
Db	315	LROSFMEFYGAGYQVEKVISHPNDYDSKTKNNIDIALMLQKPLTFENDLVKPYCLPBPNGMLQ	374
Qy	121	PEQLCWISGMCATEEKTKGTSFVLAAAVLLIETPRCNSRIVYDMLTTPAMICAGFLOGNV	180
Db	375	PEQLCWISGMCATEEKTKGTSFVLAAAVLLIETPRCNSRIVYDMLTTPAMICAGFLOGNV	434
Qy	181	DSQCGDSGGLPVLTSSKNNIIMWLIGDTSMGSGCAKAYRGGYGVGNVWFFDTWITROMRADG	238
Db	435	DSQCGDSGGLPVLTSSKNNIIMWLIGDTSMGSGCAKAYRGGYGVGNVWFFDTWITROMRADG	492

RESULT 6
US-09-895-793-895
; Sequence 895, Application US/09895793

```

: APPLICANT: Xu, Jiangchun
: APPLICANT: Dillon, Davin C.
: APPLICANT: Mitcham, Jennifer L.
: APPLICANT: Harlocker, Susan L.
: APPLICANT: Jiang, Yugu
: APPLICANT: Kalos, Michael D.
: APPLICANT: Retter, Marc W.
: APPLICANT: Stolk, John A.
: APPLICANT: Day, Craig H.
: APPLICANT: Vedyck, Thomas S.
: APPLICANT: Carter, Darrick
: APPLICANT: Li, Samuel X.
: APPLICANT: Wang, Aijun
: APPLICANT: Skelky, Yasir A.W.
: APPLICANT: Hepler, William T.
: APPLICANT: Henderson, Robert A.
: APPLICANT: Hurst, John
: APPLICANT: McNeill, Patricia D.
: APPLICANT: Houghton, Raymond L.
: APPLICANT: Vinals de Bassols, Carlota
: APPLICANT: Foy, Teresa
: APPLICANT: Fanger, Gary R.
: TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
: TITLE OF INVENTION: DIAGNOSIS OF PROSTATE CANCER
: FILE REFERENCE: 210121.534C2
: CURRENT APPLICATION NUMBER: US/09/895,793
: CURRENT FILING DATE: 2001-06-29
: NUMBER OF SEQ ID NOS: 982
: SOFTWARE: FastSeq for Windows Version 3.0
: SEQ ID NO 895
: LENGTH: 492
: TYPE: PRT
: ORGANISM: Homo sapiens
: OS-09-895-793-895

```

Query Match	98.8%	Score 1303;	DB 5;	Length 492;
Best Local Similarity	98.3%;	Pred. No. 1e-118;		
Matches 234;	Conservative 3;	Mismatches 1;	Indels 0;	Gaps 0;

```

0Y 1 RIVGESALPGAMPPOVSLHTYQNVHVGGSITTEPEWITVAHAHCYERPLNNPMMHTAFAGI 60
Db 255 RIVGESALPGAMPPOVSLHTYQNVHVGGSITTEPEWITVAHAHCYERPLNNPMMHTAFAGI 314
0Y 61 LROSEMFYAGAYOYEKVISHPNYSKTKNNDIALMLQKPLTFENDLVKPYCLPNPMMLO 120
Db 315 LROSEMFYAGAYOYKVISHPNYSKTKNNDIALMLQKPLTFENDLVKPYCLPNPMMLO 374
0Y 121 PEOLCMISGWGATEKEKGTSEVLNAAKVLLIETORCSRYDYDLITPAMICAGFLOGNV 180
Db 375 PEOLCMISGWGATEKGTSEVLNAAKVLLIETORCSRYDYDLITPAMICAGFLOGNV 434
0Y 181 DSCGGDSGGPLAYTSKNNIIMWLLIGDTSMWSGCAKAYRPGVYGVNVPFTDMYIRKMRADG 238

```

```
Db 435 DSCGDSGGPLVTSNNIMWLIGDTSWGSCKAKAYRPGVGNVWFTDWTYROMKANG 492

RESULT 7
; Sequence 895, Application US/09895814
; GENERAL INFORMATION:
; APPLICANT: Xu, Jianshun
; APPLICANT: Dillon, Devin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Harlocke, Susan L.
; APPLICANT: Jiang, Yugu
; APPLICANT: Kalos, Michael D.
; APPLICANT: Ketter, Marc W.
; APPLICANT: Stolk, John A.
; APPLICANT: Day, Craig H.
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Carter, Derrick
; APPLICANT: Li, Samuel X.
; APPLICANT: Wang, Aijun
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Hepler, William T.
; APPLICANT: Henderson, Robert A.
; APPLICANT: Hurral, John
; APPLICANT: McNeill, Patricia D.
; APPLICANT: Houghlon, Raymond L.
; APPLICANT: Vinals de Bassols, Carlota
; APPLICANT: Foy, Teresa
; APPLICANT: Fanger, Gary R.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; FILE REFERENCE: 210121.427C26
; CURRENT APPLICATION NUMBER: US/09/895, 814
; CURRENT FILING DATE: 2001-06-29
; NUMBER OF SEQ ID NOS: 990
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 895
; LENGTH: 492
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-09-895-814-895

Query Match 98.8%; Score 1303; DB 5; Length 492;
Best Local Similarity 98.3%; Pred. No. 1e-118;
Matches 234; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

OY 1 RIVGESALPGAMPWQVSLHVONVHVGGSITTPEMVITAACHVEKPLNPNMHTAFAGI 60
Db 255 RIVGESALPGAMPWQVSLHVONVHVGGSITTPEMVITAACHVEKPLNPNMHTAFAGI 314
OY 61 LRQSFMEYAGYQVEKVISHPNVDKTKKNDIALMKLQKPLTFNDLVKPVCLPFGMLQ 120
Db 315 LRQSFMEYAGYQVEKVISHPNVDKTKKNDIALMKLQKPLTFNDLVKPVCLPFGMLQ 374
OY 121 PEQLCWISGKATEEKGKTSVLAIAKYLLETQRCNSRYVDNLITPAMICAGFLQGNV 180
Db 375 PEQLCWISGKATEEKGKTSVLAIAKYLLETQRCNSRYVDNLITPAMICAGFLQGNV 434
OY 181 DSCGDSGGPLVTSNNIMWLIGDTSWGSCKAKAYRPGVGNVWFTDWTYROMKANG 238
Db 435 DSCGDSGGPLVTSNNIMWLIGDTSWGSCKAKAYRPGVGNVWFTDWTYROMKANG 492

RESULT 8
; Sequence 53, Application US/09715427
; GENERAL INFORMATION:
; APPLICANT: Quinn, Kerry E.
; APPLICANT: Spytek, Kimberly A.
; APPLICANT: Majumder, Kumud
; APPLICANT: Vernet, Corine
; APPLICANT: Herrmann, John L.

Db 435 DSCGDSGGPLVTSNNIMWLIGDTSWGSCKAKAYRPGVGNVWFTDWTYROMKANG 492

APPLICANT: Burgess, Catherine
APPLICANT: Fernandes, Elma
APPLICANT: Taupier Jr., Raymond
APPLICANT: Rastelli, Luca
APPLICANT: Curagen Corporation
TITLE OF INVENTION: NOVEL POLYPEPTIDES AND NUCLEIC ACIDS ENCODING SAME
FILE REFERENCE: 15966-598
CURRENT APPLICATION NUMBER: US/09/715,427
CURRENT FILING DATE: 2000-11-16
PRIOR APPLICATION NUMBER: U.S.S.N. 60/165,986
PRIOR FILING DATE: 1999-11-17
PRIOR APPLICATION NUMBER: U.S.S.N. 60/194,839
PRIOR FILING DATE: 2000-04-05
PRIOR APPLICATION NUMBER: U.S.S.N. 60/195,637
PRIOR FILING DATE: 2000-04-07
PRIOR APPLICATION NUMBER: U.S.S.N. 60/197,080
PRIOR FILING DATE: 2000-04-13
PRIOR APPLICATION NUMBER: U.S.S.N. 60/232,677
PRIOR FILING DATE: 2000-09-15
PRIOR APPLICATION NUMBER: U.S.S.N. 60/181,347
PRIOR FILING DATE: 2000-02-09
PRIOR APPLICATION NUMBER: U.S.S.N. 60/194,195
PRIOR FILING DATE: 2000-04-03
NUMBER OF SEQ ID NOS: 53
SOFTWARE: Patentin Ver. 2.1
SEQ ID NO 53
LENGTH: 229
TYPE: PRT
ORGANISM: Homo sapiens
US-09-715-427-53

Query Match 95.8%; Score 1263; DB 5; Length 229;
Best Local Similarity 99.1%; Pred. No. 3.1e-115;
Matches 227; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 2 IVGESALPGAMPWQVSLHVONVHVGGSITTPEMVITAACHVEKPLNPNMHTAFAGI 61
Db 1 IVGESALPGAMPWQVSLHVONVHVGGSITTPEMVITAACHVEKPLNPNMHTAFAGI 60
OY 62 ROSFMEYAGYQVEKVISHPNVDKTKKNDIALMKLQKPLTFNDLVKPVCLPFGMLQ 121
Db 61 ROSFMEYAGYQVEKVISHPNVDKTKKNDIALMKLQKPLTFNDLVKPVCLPFGMLQ 120
OY 122 EQLCWISGKATEEKGKTSVLAIAKYLLETQRCNSRYVDNLITPAMICAGFLQGNV 181
Db 121 EQLCWISGKATEEKGKTSVLAIAKYLLETQRCNSRYVDNLITPAMICAGFLQGNV 180
OY 182 SCQDSDGGPLVTSNNIMWLIGDTSWGSCKAKAYRPGVGNVWFTDWTYROMKANG 230
Db 181 SCQDSDGGPLVTSNNIMWLIGDTSWGSCKAKAYRPGVGNVWFTDWTYROMKANG 229

RESULT 9
; Sequence 3414, Application US/09760475
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT249
; CURRENT APPLICATION NUMBER: US/09/760,475
; CURRENT FILING DATE: 2001-01-16
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 4122
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 3414
; LENGTH: 163
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-09-760-475-3414

Query Match 67.8%; Score 894; DB 5; Length 163;
```

Best Local Similarity 100.0%; Pred. No. 1.6e-79;
Matches 163; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 76 KYISHPNVDSKRNNDIALMLKLOKPLTFNDLYKPVCLPMPGMLDEOJLCWISGKATEE 135

Db 1 KYISHPNVDSKRNNDIALMLKLOKPLTFNDLYKPVCLPMPGMLDEOJLCWISGKATEE 60

OY 136 KKTSEVNLAAVLIETORCSRYVDNLITPAMICAGFLGNDVSCGDSGGLVTSK 195

Db 61 KKTSEVNLAAVLIETORCSRYVDNLITPAMICAGFLGNDVSCGDSGGLVTSK 120

OY 196 NNIMWLIQDTSWGGCAKAYRPGVGNVWFTDWTYRQMRADG 238

Db 121 NNIMWLIQDTSWGGCAKAYRPGVGNVWFTDWTYRQMRADG 163

RESULT 10

US-09-925-301-1017
Sequence 1017, Application US/09925301

GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies

FILE REFERENCE: PA106
CURRENT APPLICATION NUMBER: US/09/925,301

CURRENT FILING DATE: 2001-08-10
PRIOR APPLICATION NUMBER: PCT/US00/05882

PRIOR FILING DATE: 2000-03-08
PRIOR APPLICATION NUMBER: 60/124,270

PRIOR FILING DATE: 1999-03-12
NUMBER OF SEQ ID NOS: 1694

SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 1017

LENGTH: 248
TYPE: PRT
ORGANISM: Homo sapiens

US-09-925-301-1017

Query Match 48.9%; Score 645.5; DB 5; Length 248;

Best Local Similarity 52.7%; Pred. No. 3.9e-55;
Matches 126; Conservative 30; Mismatches 78; Indels 5; Gaps 4;

OY 1 RIVGESALPGAMPQVSLHVQNVHVGSGSIITPEWIVTAHCEKPLNNPMHTAFAGI 60

Db 10 RIVGSMMSLSQMPQASLOFGYHLCGGSVITPLMITTAHCV-YDYLPRKSWITOVGL 68

OY 61 LRQSEMFYGA-GYQEKVISHPNVDSKRNNDIALMLKLOKPLTFNDLYKPVCLPMPGML 119

Db 69 V--SLIDNPAPSHLVERKIYHSHKYPKRLGNDIALMLKLAGPLTFNEMIQPVCLPNSSENF 126

OY 120 QPEOLCWSGKATEE-KKTSEVNLAAVLIETORCSRYVDNLITPAMICAGFLG 178

Db 127 PQKVCWTSQMGKATEDEGADASVPLNHAAPPLISNKRICHNRDVGGLISPSMLCAGYLIG 186

OY 179 NVDSGCGDSGGLVTSKNNIMWLIQDTSWGGCAKAYRPGVGNVWFTDWTYRQMRAD 237

Db 187 GVDSCGDSGGLVTCGERRLMKLVGATSFQIGCAEVNKRGCVTYRVTSLFDWIHEQMRD 245

RESULT 11

PCT-US01-18569-3126
Sequence 3126, Application PC/TUS0118569

GENERAL INFORMATION:
APPLICANT: Human Genome Sciences, Inc.
TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies

FILE REFERENCE: PA133PCT
CURRENT APPLICATION NUMBER: PCT/US01/18569

CURRENT FILING DATE: 2001-06-07
PRIOR APPLICATION NUMBER: 60/209,467

PRIOR FILING DATE: 2000-06-07
NUMBER OF SEQ ID NOS: 4360

SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 3126

LENGTH: 394

TYPE: PRT
ORGANISM: Homo sapiens
PCT-US01-18569-3126

Query Match

Best Local Similarity 48.9%; Score 645.5; DB 1; Length 394;

Matches 126; Conservative 30; Mismatches 78; Indels 5; Gaps 4;

OY 1 RIVGESALPGAMPQVSLHVQNVHVGSGSIITPEWIVTAHCEKPLNNPMHTAFAGI 60

Db 156 RIVGSMMSLSQMPQASLOFGYHLCGGSVITPLMITTAHCV-YDYLPRKSWITOVGL 214

OY 61 LRQSEMFYGA-GYQEKVISHPNVDSKRNNDIALMLKLOKPLTFNDLYKPVCLPMPGML 119

Db 215 V--SLIDNPAPSHLVERKIYHSHKYPKRLGNDIALMLKLAGPLTFNEMIQPVCLPNSSENF 272

OY 120 QPEOLCWSGKATEE-KKTSEVNLAAVLIETORCSRYVDNLITPAMICAGFLG 178

Db 273 PQKVCWTSQMGKATEDEGADASVPLNHAAPPLISNKRICHNRDVGGLISPSMLCAGYLIG 332

OY 179 NVDSGCGDSGGLVTSKNNIMWLIQDTSWGGCAKAYRPGVGNVWFTDWTYRQMRAD 237

Db 333 GVDSCGDSGGLVTCGERRLMKLVGATSFQIGCAEVNKRGCVTYRVTSLFDWIHEQMRD 391

RESULT 12

US-09-715-427-15
Sequence 15, Application US/09715427

GENERAL INFORMATION:
APPLICANT: Quinn, Kerry E.

APPLICANT: Spytek, Kimberly A.

APPLICANT: Majumder, Kumud

APPLICANT: Verneil, Corine

APPLICANT: Herrmann, John L.

APPLICANT: Burgess, Catherine

APPLICANT: Fernandes, Elma

APPLICANT: Taupier Jr., Raymond

APPLICANT: Rastelli, Luca

APPLICANT: Curagen Corporation

TITLE OF INVENTION: NOVEL POLYPEPTIDES AND NUCLEIC ACIDS ENCODING SAME

FILE REFERENCE: 15966-598

CURRENT APPLICATION NUMBER: US/09/715,427

PRIOR FILING DATE: 2000-11-16

PRIOR APPLICATION NUMBER: U.S.S.N. 60/165,986

PRIOR FILING DATE: 1999-11-17

PRIOR APPLICATION NUMBER: U.S.S.N. 60/194,839

PRIOR FILING DATE: 2000-04-05

PRIOR APPLICATION NUMBER: U.S.S.N. 60/195,637

PRIOR FILING DATE: 2000-04-07

PRIOR APPLICATION NUMBER: U.S.S.N. 60/197,080

PRIOR FILING DATE: 2000-04-13

PRIOR APPLICATION NUMBER: U.S.S.N. 60/232,677

PRIOR FILING DATE: 2000-09-15

PRIOR APPLICATION NUMBER: U.S.S.N. 60/181,347

PRIOR FILING DATE: 2000-02-09

PRIOR APPLICATION NUMBER: U.S.S.N. 60/194,195

PRIOR FILING DATE: 2000-04-03

NUMBER OF SEQ ID NOS: 53

SOFTWARE: PatentIn Ver. 2.1

SEQ ID NO 15

LENGTH: 421

TYPE: PRT

ORGANISM: Homo sapiens

US-09-715-427-15

Query Match

Best Local Similarity 42.5%; Score 561; DB 5; Length 421;

Matches 109; Conservative 36; Mismatches 87; Indels 10; Gaps 4;

OY 1 RIVGESALPGAMPQVSLHVQNVHVGSGSIITPEWIVTAHCV---EKLNNPMHTA 56

```

Db 179 RIVGALSDSKMPQVSLHFGTTHIGCTTLDAQWVLTAAACFEVTRBKVEG--WKV 235
QY 57 FAGIRQSFMEYAGYQVEKYVISHPNYDSKTKNNIDIAIMKLOKPLTFNDLVPCVCLPMPG 116
Db 236 YAGTSMNLQLEBA--STAETIINSNYDEDEDDYDIALMRSLKPLTSAHHPACLPMPHG 293
QY 117 MMLQPEQLCMTISGWGATEE-KGTSEVLNAKVLLEIETORNSRYVNDLITPAMICAGF 175
Db 294 QTFSLNETCMTGTFGKTRTETDDKTSPLREVOVNLIDFKKCDYLVDSYLTPLRMKACAD 353
QY 176 LOGNVDSGCGSGGPLYSKNNIMWLIGDTSWGSCKAKAYRPGVGNMVFETDWTYRQMR 235
Db 354 LRGGDSGCGSGGPLYVCEQNNRMVLAGVTSWGTGCGGRNRPVYTKVTEVLPMTYSKME 413
QY 236 AD 237
Db 414 SE 415

```

```

RESULT 13
US-09-875-195-922
: Sequence 922, Application US/09875195
: GENERAL INFORMATION:
: APPLICANT: Rosen et al.
: TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
: FILE REFERENCE: PA132N2
: CURRENT APPLICATION NUMBER: US/09/875,195
: CURRENT FILING DATE: 2001-06-07
: Prior application data removed - consult PAM or file wrapper
: NUMBER OF SEQ ID NOS: 922
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 922
: LENGTH: 343
: TYPE: PRT
: ORGANISM: Homo sapiens
: FEATURE:
: NAME/KEY: SITE
: LOCATION: (309)
: OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-875-195-922

```

```

Query Match 40.1%; Score 529.5; DB 5; Length 343;
Best Local Similarity 41.2%; Pred. No. 1,le-43;
Matches 101; Conservative 41; Mismatches 92; Indels 11; Gaps 3;

QY 1 RIVGESALPGAMPQVSLHVNHYVCGSITTEPMITVAHCVKPLNPMHMTAFAGI 60
Db 88 RIVGGRDTSIGRMPQVSLRYDGAHLGSGSLSGDMVLTAAHCFERRRVLSRMVRFAGA 147
QY 61 LRQSFMEYAGYQVEKYVISHPNY-----DSKTKNNIDIAIMKLOKPLTFNDLVPCVCLP 114
Db 148 VAQASP-HGLQGVQAVVYHGGYLPFRDPNSEBNSNDIALVHLSPLPLEYIOPVCLPAA 206
QY 115 PGMALQPEQLCMTISGWGATEEKTSEVLNAKVLLEIETORNSRYVNDLITPAMICAG 174
Db 207 AGQALVDGKICVTGWTGYGQAGVLAQEARVPLISNDVCGADFGNIGKPKMFCAG 266
QY 175 FLQGVNDSGCGSGGPLY-----TSKNNIMWLIGDTSWGSCKAKAYRPGVGNMVFETDWT 230
Db 267 YPEGIDACGDSGSPFYCEDSISRTPRMRLCGIVSWGTGCAKXAKOPGVYTKVSDPREMI 326
QY 231 YRQMR 235
Db 327 FOAIR 331

```

```

RESULT 14
US-09-410-362B-6
: Sequence 6, Application US/09410362B
: GENERAL INFORMATION:
: APPLICANT: THE REGENTS OF THE UNIVERSITY OF CALIFORNIA

```

```

: APPLICANT: Craik, Charles
: APPLICANT: Takeuchi, Toshihiko
: TITLE OF INVENTION: MEMBRANE TYPE SERINE PROTEASE 1 (MT-SP10 AND USES THEREOF)
: FILE REFERENCE: 305T-9001-000S
: CURRENT APPLICATION NUMBER: US/09/410,362B
: CURRENT FILING DATE: 1999-09-30
: NUMBER OF SEQ ID NOS: 33
: SOFTWARE: PatentIn version 3.0
: SEQ ID NO 6
: LENGTH: 255
: TYPE: PRT
: ORGANISM: Artificial Sequence
: FEATURE:
: OTHER INFORMATION: Protein fragment/domain
US-09-410-362B-6

```

```

Query Match 39.7%; Score 523.5; DB 5; Length 255;
Best Local Similarity 41.0%; Pred. No. 3e-43;
Matches 100; Conservative 41; Mismatches 92; Indels 11; Gaps 3;

QY 2 IYGGESALPGAMPQVSLHVNHYVCGSITTEPMITVAHCVKPLNPMHMTAFAGI 61
Db 1 IYGGGRDTSIGRMPQVSLRYDGAHLGSGSLSGDMVLTAAHCFERRRVLSRMVRFAGA 60
QY 62 ROSEMEYAGYQVEKYVISHPNY-----DSKTKNNIDIAIMKLOKPLTFNDLVPCVCLP 115
Db 61 AQASP-HGLQGVQAVVYHGGYLPFRDPNSEBNSNDIALVHLSPLPLEYIOPVCLPAA 119
QY 116 GMLQPEQLCMTISGWGATEEKTSEVLNAKVLLEIETORNSRYVNDLITPAMICAGF 175
Db 120 GOALVDGKICVTGWTGYGQAGVLAQEARVPLISNDVCGADFGYQNIKPKMFCAGY 179
QY 176 LOGNVDSGCGSGGPLY-----TSKNNIMWLIGDTSWGSCKAKAYRPGVGNMVFETDWT 231
Db 180 PEGIDACGDSGSPFYCEDSISRTPRMRLCGIVSWGTGCAKXAKOPGVYTKVSDPREMI 239
QY 232 RQMR 235
Db 240 QAIR 243

```

```

RESULT 15
PCT-US01-18568-2
: Sequence 2, Application PC/TUS0118568
: GENERAL INFORMATION:
: APPLICANT: Darrow, Andrew L
: APPLICANT: Q1, Jain-shen
: TITLE OF INVENTION: DNA encoding human serine protease D-G
: FILE REFERENCE: ORT-1273
: CURRENT APPLICATION NUMBER: PCT/US01/18568
: CURRENT FILING DATE: 2001-06-08
: NUMBER OF SEQ ID NOS: 9
: SOFTWARE: PatentIn Ver. 2.1
: SEQ ID NO 2
: LENGTH: 435
: TYPE: PRT
: ORGANISM: Homo sapiens
PCT-US01-18568-2

```

```

Query Match 39.5%; Score 521; DB 1; Length 435;
Best Local Similarity 44.6%; Pred. No. 1e-42;
Matches 108; Conservative 36; Mismatches 84; Indels 14; Gaps 7;

```

```

QY 1 RIVGESALPGAMPQVSLHVNHYVCGSITTEPMITVAHCVKPLNPMHMTAFAGI 60
Db 202 RIVGGEASVDSMPQVSIQYDKOHVCGSITLDPHVVLTAAHCFERKH-TDVFNKRVAGS 260
QY 61 LR-QSFMEYAGYQVEKYI---SHPNYDSKTKNNIDIAIMKLOKPLTFNDLVPCVCLPMPG 116

```

```

Db 261 DKIGSF-----PSLAVAKIIIEFNPMY---PKNDIALMKLOPILFSGTVRPICLPFPD 313
QY 117 MMLQPEQLCMISGMGATEEK-GKTSFVLAARVLLIEFORCNSRYVYDNLITPAMICAGF 175
Db 314 EELTPATPLMILIGMGFTKONGKMSDILQASVOYIDSTRCNADDAVOGEVTEKMCAGI 373
QY 176 LOGNVDSGCGDSGGPLVTSKNINIMWLIGDTSWGSCKAKARPGVYGNVAVFTDVIYRQMR 235
Db 374 PEGGYDTGCGDSGGPLMYSDQ-WHVYGLVSMGCGGPGSTPGVYTKVSATILNMIYNMK 432
QY 236 AD 237
Db 433 AE 434

```

Search completed: September 26, 2001, 09:23:23
 Job time: 138 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 26, 2001, 09:21:05 : Search time 20.35 Seconds
(without alignments)
497.811 Million cell updates/sec

Title: US-09-615-285-2
2717
Sequence: 1 MALNCGSPPAIGPYENHGX.....VYGNVWFETDVIYROMRADG 492

Scoring table: BIOSUM62
Gapop 10.0, Gapext 0.5

Searched: 197339 seqs, 20590346 residues

Total number of hits satisfying chosen parameters: 197339

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/1aa/5A.COMB.pep:*
2: /cgn2_6/ptodata/2/1aa/5B.COMB.pep:*
3: /cgn2_6/ptodata/2/1aa/6A.COMB.pep:*
4: /cgn2_6/ptodata/2/1aa/6B.COMB.pep:*
5: /cgn2_6/ptodata/2/1aa/PTCUS.COMB.pep:*
6: /cgn2_6/ptodata/2/1aa/Backfile1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2717	100.0	492	US-09-342-749-2	Sequence 2, App11
2	1540	56.7	283	US-08-807-151-1	Sequence 1, App11
3	676.5	24.9	435	US-09-008-271A-6	Sequence 6, App11
4	660	24.3	798	US-08-200-900A-2	Sequence 2, App11
5	660	24.3	798	PCR-US94-00616-2	Sequence 2, App11
6	558.5	20.6	855	US-09-027-337-2	Sequence 2, App11
7	556	20.5	638	US-08-681-151-3	Sequence 3, App11
8	553.5	20.4	248	US-08-944-483-63	Sequence 63, App11
9	539.5	19.9	416	US-09-000-846-2	Sequence 1, App11
10	534	19.7	356	US-08-681-151-1	Sequence 1, App11
11	528.5	19.5	256	US-09-027-337-3	Sequence 3, App11
12	527.5	19.4	418	US-08-508-448C-25	Sequence 25, App11
13	524	19.3	238	US-08-944-483-64	Sequence 64, App11
14	523.5	19.3	235	US-08-944-483-67	Sequence 67, App11
15	521.5	19.2	235	US-08-807-151-3	Sequence 3, App11
16	518.5	19.1	812	US-08-248-629A-1	Sequence 1, App11
17	518.5	19.1	812	US-08-451-933-1	Sequence 1, App11
18	518.5	19.1	812	US-08-452-260-1	Sequence 1, App11
19	518.5	19.1	812	US-08-326-785-1	Sequence 1, App11
20	518.5	19.1	812	US-08-612-788-1	Sequence 1, App11
21	518.5	19.1	812	US-08-605-598B-1	Sequence 1, App11
22	518.5	19.1	812	US-08-429-743-1	Sequence 1, App11
23	518.5	19.1	812	US-08-866-735-1	Sequence 1, App11
24	518.5	19.1	812	US-09-066-028-1	Sequence 1, App11
25	518.5	19.1	812	PCR-US95-05107-1	Sequence 1, App11
26	513.5	18.9	232	US-08-508-448C-19	Sequence 19, App11
27	511.5	18.8	235	US-08-944-483-65	Sequence 65, App11

28	495.5	18.2	250	US-08-944-483-68	Sequence 68, App11
29	459.5	16.9	314	US-09-008-271A-3	Sequence 3, App11
30	452	16.6	546	5200340-6	Patent No. 5200340
31	450	16.6	270	US-08-978-404B-8	Sequence 8, App11
32	449	16.5	655	US-08-148-910-12	Sequence 12, App11
33	449	16.5	655	US-08-448-937A-12	Sequence 12, App11
34	447	16.5	276	US-09-016-366A-15	Sequence 15, App11
35	447	16.5	276	US-08-978-404B-21	Sequence 21, App11
36	446	16.4	230	US-08-379-621-2	Sequence 2, App11
37	446	16.4	230	US-08-147-000B-2	Sequence 2, App11
38	446	16.4	230	US-08-889-078-2	Sequence 2, App11
39	445.5	16.4	791	US-08-643-219-1	Sequence 2, App11
40	445.5	16.4	791	US-09-131-995-1	Sequence 1, App11
41	445.5	16.4	791	US-08-832-087B-1	Sequence 1, App11
42	445.5	16.4	791	US-08-851-350-1	Sequence 1, App11
43	445.5	16.4	791	US-09-132-154-1	Sequence 1, App11
44	445.5	16.4	810	US-07-854-603-2	Sequence 2, App11
45	445.5	16.4	810	US-08-147-000B-29	Sequence 29, App11

ALIGNMENTS

RESULT 1
US-09-342-749-2
; Sequence 2, Application US/09342749
; Patent No. 5166194
; GENERAL INFORMATION:
; APPLICANT: Wong, Alexander K.C.
; APPLICANT: Tavligian, Sean V.
; APPLICANT: Teng, David H.-F.
; APPLICANT: Myriad Genetics, Inc.
; TITLE OF INVENTION: TMRSS2 is a Tumor Suppressor
; FILE REFERENCE: 2318-202
; CURRENT APPLICATION NUMBER: US/09/342,749
; EARLIER FILING DATE: 1999-06-29
; EARLIER APPLICATION NUMBER: US 60/091,044
; EARLIER FILING DATE: 1998-06-29
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO: 2
; LENGTH: 492
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-342-749-2

Query Match 100.0%; Score 2717; DB 4; Length 492;
Best Local Similarity 100.0%; Pred. No. 2.1e-241;
Matches 492; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	MALNCGSPPAIGPYENHGXQPNPYPAPQPTVYVYEVHAPQYSPVQYAPRVLQA	60
DB	1	MALNCGSPPAIGPYENHGXQPNPYPAPQPTVYVYEVHAPQYSPVQYAPRVLQA	60
QY	61	SNPVCTOPSPSGYCTSTKTKALCITTLTGFLGAAALAKLKKPMKSCNSGIEC	120
DB	61	SNPVCTOPSPSGYCTSTKTKALCITTLTGFLGAAALAKLKKPMKSCNSGIEC	120
QY	121	DSSGTCINPNCMDGVSHCPGEGDENRCVLYGNPFILOYSSQKSMHPVCCDDNNENY	180
DB	121	DSSGTCINPNCMDGVSHCPGEGDENRCVLYGNPFILOYSSQKSMHPVCCDDNNENY	180
QY	181	GRACADMDGKKNFYSSQGIYVDSGSTSPFKLTSAGNVDIYKKLYHSDACSKAVYSLR	240
DB	181	GRACADMDGKKNFYSSQGIYVDSGSTSPFKLTSAGNVDIYKKLYHSDACSKAVYSLR	240
QY	241	CACGYNLNSRSRIVGGSALPGAMPQVSLHYQNVHVCSSITTPETVYAAHCYER	300
DB	241	CACGYNLNSRSRIVGGSALPGAMPQVSLHYQNVHVCSSITTPETVYAAHCYER	300
QY	301	PLNNPHTAFACILROSEFYGAGYQVERKISHPYDSTKKNNDIALKLQPLTFNDL	360
DB	301	PLNNPHTAFACILROSEFYGAGYQVERKISHPYDSTKKNNDIALKLQPLTFNDL	360

Db	301	PLNNPMHWTATFAGILRQSFMEFYAGYQVEKVISHPNPDSTKKNNDIALMKLQPLTFENDL	360
Qy	361	VKPCPLPBPCHMLQPEQLCWTISGHCATDEEKGTSEVLNNAKVLLETRQNSRYVDNLI	420
Db	361	VKPCPLPBPCHMLQPEQLCWTISGHCATDEEKGTSEVLNNAKVLLETRQNSRYVDNLI	420
Qy	421	TPAMICAGFLOGNVDSCGGSGPVLNLSKNNIMWLIDDTISGSCCAAYRPQVYGNMVF	480
Db	421	TPAMICAGFLOGNVDSCGGSGPVLNLSKNNIMWLIDDTISGSCCAAYRPQVYGNMVF	480
Qy	481	TDWIYRQMRADG	492
Db	481	TDWIYRQMRADG	492

RESULT 2
US-08-807-151-1
: Sequence 1, Application US/08807151
: Datafile US/08807151

```

1      TITLE OF INVENTION:  NOVEL HUMAN PROSTATE-ASSOCIATED
2      TITLE OF INVENTION:  PROTEASE
3      NUMBER OF SEQUENCES:  5
4      CORRESPONDENCE ADDRESS:
5      ADDRESSEE:  Incyte Pharmaceuticals, Inc.
6      STREET:  3174 Porter Drive
7      CITY:  Palo Alto
8      STATE:  CA
9      COUNTRY:  US
10     ZIP:  94304
11     COMPUTER READABLE FORM:
12     MEDIUM TYPE:  Diskette
13     COMPUTER:  IBM compatible
14     OPERATING SYSTEM:  DOS
15     SOFTWARE:  FASTED Version 2.0
16     CURRENT APPLICATION DATA:
17     APPLICATION NUMBER:  US/08/807,151
18     FILING DATE:  Filed Herewith
19     CLASSIFICATION:  424
20     PRIOR APPLICATION DATA:
21     APPLICATION NUMBER:
22     FILING DATE:
23     ATTORNEY/AGENT INFORMATION:
24     NAME:  Billings, Lucy J.
25     REGISTRATION NUMBER:  36,749
26     REFERENCE/DOCKET NUMBER:  PF-0227 US
27     TELECOMMUNICATION INFORMATION:
28     TELEPHONE:  415-845-0555
29     TELEFAX:  415-845-4166
30     INFORMATION FOR SEQ ID NO:  1:
31     SEQUENCE CHARACTERISTICS:
32     LENGTH:  283 amino acids
33     TYPE:  amino acid
34     STRANDEDNESS:  single
35     TOPOLOGY:  linear
36     IMMEDIATE SOURCE:
37     LIBRARY:  SCORNO701
38     CLONE:  556016
39     JS-08-807-151-1

```

Query Match	56.78;	Score 1540;	DB 3;	length 283;
Best Local Similarity	99.68;	Pred. No. 1.3e-133;		
Matches 282; Conservative	0;	Mismatches 1;	Indels 0;	Gaps 0;

Qy 210 MKLTSAGNVNDIYKKLYHSDACSSKAAVYSLRCLACIAGVNLNLSRQSRIVGESALPGAMP 269

Dy 1 MKLTSAGNVNDIYKKLYHSDACSSKAAVYSLRCLACIAGVNLNLSRQSRIVGESALPGAMP 60

Qy 270 QVSLHVQNVNHCSSGSIITPBEITYAAACVEKPLNNPMMHAFAGILRQSMFPGAGIQVE 329

Db	61	QVSLHVGAVHVGCGSIIIPREIVTAANHCVEKPLNNPIHMTAFAGILNQSMFAGAGQVE	120
QY	330	KVISHPNTDSTKKNDDIALMKLQPLPFPNDLVKPVCLPDPGMILQPEQLCISGKGATEE	389
Db	121	KVISHPNDSTKKNDDIALMKLQPLPFPNDLVKPVCLPDPGMILQPEQLCISGKGATEE	180
QY	330	KGKTSVYLNAAKVLIIETQRNSRYVDNLITPMTAGFLQGVNDSGODSGGLVYTSK	449
Db	181	KGKTSVYLNAAKVLIIETQRNSRYVDNLITPMTAGFLQGVNDSGODSGGLVYTSK	240
QY	450	NNIMWLIGDTSMSGCAKAYRPGYGVNMVETDITROMRADG	492
Db	241	NNIMWLIGDTSMSGCAKAYRPGYGVNMVETDITROMRADG	283

RESULT 3
US-09-008-271A-6
; Sequence 6, Application US/09008271A
Result 3: 000070

```

1 Hillman, Jennifer L.
2 Yue, Henry
3 Guegler, Karl J.
4 Corley, Neil C.
5 Tang, Tom Y.
6 Shah, Purvi
7
8 TITLE OF INVENTION: HUMAN PROTEASE MOLECULES
9
10 NUMBER OF SEQUENCES: 24
11
12 CORRESPONDENCE ADDRESS:
13 ADDRESS: Incyte Pharmaceuticals, Inc.
14 STREET: 3174 Porter Dr.
15
16 CITY: Palo Alto
17
18 STATE: CA
19
20 COUNTRY: USA
21
22 ZIP: 94304
23
24 COMPUTER READABLE FORM:
25 MEDIUM TYPE: Diskette
26
27 COMPUTER: IBM Compatible
28
29 OPERATING SYSTEM: DOS
30
31 SOFTWARE: FASTSEQ for Windows Version 2.0.
32
33 CURRENT APPLICATION DATA:
34 APPLICATION NUMBER: US/09/008,771A
35
36 FILING DATE: 16-Jan-1998
37
38 PRIOR APPLICATION DATA:
39 APPLICATION NUMBER: <unknown>
40
41 FILING DATE: <unknown>
42
43 ATTORNEY/AGENT INFORMATION:
44 NAME: Mohan-Peterson, Sheela
45 REGISTRATION NUMBER: 41,201
46 REFERENCE/DOCKET NUMBER: PF-0458 US
47
48 TELECOMMUNICATION INFORMATION:
49 TELEPHONE: 650-855-0555
50
51 TELEFAX: 650-845-4166
52
53 INFORMATION FOR SEQ ID NO: 6:
54
55 SEQUENCE CHARACTERISTICS:
56
57 LENGTH: 435 amino acids
58
59 TYPE: amino acid
60
61 STRANDEDNESS: single
62
63 TOPOLOGY: linear
64
65 IMMEDIATE SOURCE:
66
67 LIBRARY: COLNOT13
68
69 CLONE: 1337018
70
71 SEQUENCE DESCRIPTION: SEQ ID NO: 6 :
72
73 US-09-008-771A-6

```

```

;          SEQUENCE DESCRIPTION: SEQ ID NO: 6 :
US-09-008-271A-6

```

Query Match	24.9%	Score 676.5;	DB 4;	Length 435;
Best Local Similarity	39.1%;	Pred. NO. 5e-54;		
Matches 150;	Conservative 57;	Mismatches 128;	Indels 49;	Gaps 13;
133	CDGVSHCCGGEDENCCVRLY--GP-----NFILOYISSQKRSMPHVCDDMMNNTG	181		

```

Db 75 CDGELDCPIGDEDEHCVKSPFEGPAVAVRLSKRSTLQVLDSTATGNWFSACFDNFTAL 134
QY 182 RAACRDMGY--KNNFYSSQ-----GIYDSSGSTFEMKIMTSAGNVDIYKKLYSDACS 232
Db 135 ETACRQMKYSSKPTFRRAEIGPPDDLDVETENSOELMRNRSQ-----PCL 182
QY 233 SKAVYSLICAGVALNLSRQSIYIGESALPGAMPQVSLHYOVNHYVCGGSIITPEWY 292
Db 183 SGLVSLHCLACGESL---KTPRYVGGEEASVDSMPQVSYQYDQYHCGGSIIDPHVYL 239
QY 293 TAAHCVEKPLNPMHMTAFAGILR-QSFMEYAGYQVEKVI---SHPNYDSKTKNNDIAL 348
Db 240 TAAHCFRKH--TDVFNWKYRAGSDKLGSE---PSLAVAKIITIEFPNMY---PKNDIAL 291
QY 349 MKLQKPLTFNDLVKPVCLPNPGMMLQPBOLCISGMGATEEK-GKTSVYLAANKYLLIET 407
Db 292 MKLQPLTFSTGVPRICLPFEDEELTPATPLMIGMFTKQGGMSDILLQASVOIDS 351
QY 408 ORCNSRYVDNLITPAMICAGFLOGNVDSGQSGPLVTSKNNIMWLGDTSMGSCAK 467
Db 352 TRCNDADYVQGEVEYKMKACAGIPGCGVDTCCGDSGGLMTQSDQ-WHYVYGIYSMGYGGG 410
QY 468 AYRGEVGNVMTDWTIVYRQMRAD 491
Db 411 PSTPGVYTKVSAYLMIYVWKAE 434

RESULT 4
US-08-200-900A-2
; Sequence 2, Application US/08200900A
; Patent No. 5655566
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: CLONING OF ENTEROKINASE AND METHOD OF USE
; NUMBER OF SEQUENCES: 38
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc. - Legal Affairs
; STREET: 87 Cambridgepark Drive
; CITY: Cambridge
; STATE: MA
; COUNTRY: USA
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/200,900A
; FILING DATE: 23-FEB-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Melner, Maureen C.
; REGISTRATION NUMBER: 31,544
; REFERENCE/DOCKET NUMBER: GI 5201-FWC
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 876-1170 X8574
; TELEFAX: (617) 876-5851
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 798 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-200-900A-2

Query Match 24.3%; Score 660; DB 1; Length 798;
Best Local Similarity 33.9%; Pred. No. 3.9e-52;
Matches 150; Conservative 67; Mismatches 180; Indels 46; Gaps 12;
QY 65 VCTOPKSPSGVCTSKTKALCITITL-----GTFLVGALALAGLMLKFMGSKCSNSG 117

```

```

Db 374 VYTGCPVNDVSTTNRMVTLFTIDNMLAKQCFKANFTGYGLG-----IPECKEDN 426
QY 118 IECDSSTGTCINSNMCDGVSHCPGGEDEKRCYRLYG-----PNFIIQVYSSORKSMH 169
Db 427 FQC-KDGEICPLVNLCDGPPHCKDGSDEAHCVRLNGTTDSSGLVQFRIO-----SIWH 479
QY 170 PVCODDMWENGRACRDMGYKNNFYSSQGIYDSSGSTSFEMKIMTSAGNVDIYKKLYSD 229
Db 480 VACAENWTTQISDVCQLLGLGTG--NSSVPFTSTGGCPYVNLNAPNCSLI---LTPSQ 534
QY 230 ACSSRAVYSLRC--IACGVNLSRSQSIYIGESALPGAMPQVSLHYOVNHYVCGGSIIT 286
Db 535 QCLEDSLILILQCNYSKCGKLVTEVSPRIYVSGDSRSGAMWVVALFDDQVCGASLY 594
QY 287 TPFWITVTAHCVEKPLNPMHMTAFAGILRQSFMEYAGYQVEKVI---KVISHPNYDSKT 341
Db 595 SRDWLVSAHCVYGRNMEPSKWKAVLGLMASNL---TSPQIETRLIDQIVINPNYKRR 651
QY 342 KNDIALMKLQKPLTFNDLVKPVCLPNPGMMLQPBOLCISGMGATEEKKTSEVYLAANK 401
Db 652 KNDIALMMHLEKKNVITTDIOPICLPEENQVFPFGRTCSIAGMALIYQGSTADYVQED 711
QY 402 VLIETORCNSRYVDNLITPAMICAGFLOGNVDSGQSGPLVTSKNNIMWLGDTSM 461
Db 712 VPLLENKQOQOMPEYN-ITENMVCAGYVAGGVDSGQSGGLMCGENRMVLAGVTSF 770
QY 462 GSGCAKATRPQYGVNMTDWTI 484
Db 771 GYOCALPNRPGYARVPRTETWI 793

RESULT 5
PCT-US94-00616-2
; Sequence 2, Application PC/TUS9400616
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: CLONING OF ENTEROKINASE AND METHOD OF USE
; NUMBER OF SEQUENCES: 33
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US94/00616
; FILING DATE:
; CLASSIFICATION:
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 798 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; PCT-US94-00616-2

Query Match 24.3%; Score 660; DB 5; Length 798;
Best Local Similarity 33.9%; Pred. No. 3.9e-52;
Matches 150; Conservative 67; Mismatches 180; Indels 46; Gaps 12;
QY 65 VCTOPKSPSGVCTSKTKALCITITL-----GTFLVGALALAGLMLKFMGSKCSNSG 117
Db 374 VYTGCPVNDVSTTNRMVTLFTIDNMLAKQCFKANFTGYGLG-----IPECKEDN 426
QY 118 IECDSSTGTCINSNMCDGVSHCPGGEDEKRCYRLYG-----PNFIIQVYSSORKSMH 169
Db 427 FQC-KDGEICPLVNLCDGPPHCKDGSDEAHCVRLNGTTDSSGLVQFRIO-----SIWH 479
QY 170 PVCODDMWENGRACRDMGYKNNFYSSQGIYDSSGSTSFEMKIMTSAGNVDIYKKLYSD 229
Db 480 VACAENWTTQISDVCQLLGLGTG--NSSVPFTSTGGCPYVNLNAPNCSLI---LTPSQ 534

```



```

OY      256  IVGGESALPGAMPWQVSLHYQ---NVHVCGGSIITPEWITATAHCEV-KPLNPNWHWTAF 311
      |||| :: | ||||| | : ||||| : ||||| : || : |

```

QY 137 SHCPGDEENRCVRLYGPNFILQVYSSQQRKSWHPVCQDDNNENYGRACRDMGYKNFYS 196

```

Db 41 ILLGSDQEPYVQVQLSPDSDRLAVFDKTEGTRLLCSSRSNARVAGLCEEMGFRLALAH 100
QY 197 SGGIVDDSGSTSFMKLTSA-----GNVDIYKKLYHSDA---CSSKAVSLRCLACGVN 247
Db 101 SELDVRRTAGAN-----GTSGFECVDEGGLRLAQLRLDVIYVDCDGRGRFLTITCODCG-- 153
QY 248 LNSSRQSIIVGCEGSLPGAMPQVSLHVNQVHVGSGSIITPEWITAAHCVEKPLNNPMH 307
Db 154 RRKLPLVDRIYVGQDSISLGRMPQVSLRYDGTHLGCGSLISGDMVLTAAHCPEERRRVLSR 213
QY 308 WTAFAGLRQSFMEYAGAYQVEKYVSHPNY-----DSKTKNNDIALKLOKPLTFNDLY 361
Db 214 WRVFGAIVARTSP-HAVDLGVQAVIYHGGYLPFRDPTIDENSNDIALVHLSSSLPLEYI 272
QY 362 KPVCLPNEGMLQPEQLCWMISGMATEEKGKTSVYLNAAKYLLETQRCNSRHYVDNLIT 421
Db 273 QPVCLPAAGQALVYDGKCTVYGMWNTQFYGGQAMVLOEARPILISNEVCSNDFEGNQIK 332
QY 422 PAMICAGFLQGVNDSGCGSGPLV----TSKNINIMWLIQDTSWGSCKAKAVPGVYGVN 477
Db 333 PKMCAGYPEGGIDACQDSGSGPFVCEDSISGTSNMRICGIYVSGTCALARKPGVYTKV 392
QY 478 MVTDMYRQMR 489
Db 393 TDEREMITKAIK 404

```

```

RESULT 10
US-08-681-151-1
; Sequence 1, Application US/08681151
; Patent No. 5869637
; GENERAL INFORMATION:
; APPLICANT: Au-Young, Janice
; APPLICANT: Bandman, Olga
; APPLICANT: Braxton, Scott Michael
; APPLICANT: Goll, Surya
; TITLE OF INVENTION: A NOVEL HUMAN KALLIKREIN
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: US
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ Version 1.5
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/681,151
; FILING DATE: Herewith
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0074US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; TELEX:
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 356 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide

```

```

; IMMEDIATE SOURCE:
; LIBRARY: HEARNOT01
; CLONE: 307474
; US-08-681-151-1

```

```

Query Match 19.7%; Score 534; DB 2; Length 356;
Best Local Similarity 35.4%; Pred. No. 4.6e-41;
Matches 120; Conservative 53; Mismatches 132; Indels 34; Gaps 11;

```

```

QY 169 HPVCCDDNNENNGRAACQDMGKKNFYSSQGIYDDSGSTSFEMKLTNT---SAGNDVIYKKL 225
Db 25 HHVCADGQWELISQLCQMGIGEPSYTKL-IQDEKPERWLTLSNBSLNGTTLHLELL 83
QY 226 YHSDACSSKAVVSLRCLIA--CGVNLNSSRQSIIVGCEGSLPGAMPQVSLHVO--NVHVC 282
Db 84 VNGGCEESRSKISLCTQKQDCGXRPAAKMKRIIGGRISRGRRMPQVSLSEPSGHICG 143
QY 283 GSITPEWITVIAHCFE-----KPLNPMHMTAFAGILRQSFMEYAGAYQVEKY 331
Db 144 CVLLAKKMWLTVAHCFEGRENAAYKKVVLGINNLDHPSPV--MOTRP-----VKTI 192
QY 332 ISHPYDSKTKNNDIALKLOKPLTFNDLVKPVCLPNEGMLQPEQLCWMISGMATEEKG 391
Db 193 ILHPRTSRAVVDYDITSELSEDISETGIVRVCPLPNEGMLPEPTVYITGMG--HMGN 250
QY 392 KTSEVLNAAKYLLETQRCNSRHYVDNLITPAMICAGFLQGVNDSGCGSGPLVTSK-N 450
Db 251 KMPFKLOGEYRIISLEHCQS-YFDMKTIITRMICAGYESGVYSCMDSGGLYCEKRG 309
QY 451 NIMWLIQDTSWGSCK-AKAYRPGVYGVNMTDMYRQMR 488
Db 310 GRWTLFGLTWSGVCFSKVLPGVYVSNYSYFVEMIKROI 348

```

```

RESULT 11
US-09-027-337-3
; Sequence 3, Application US/09027337B
; Patent No. 5972616
; GENERAL INFORMATION:
; APPLICANT: O'Brien, Timothy J.
; APPLICANT: Tani moto, Hiroto shi
; TITLE OF INVENTION: Breast and Ovarian Carcinomas
; FILE REFERENCE: D6064
; CURRENT APPLICATION NUMBER: US/09/027,337B
; CURRENT FILING DATE: 1998-02-20
; NUMBER OF SEQ ID NOS: 13
; SEQ ID NO 3
; LENGTH: 256
; TYPE: PRT
; ORGANISM: Unknown
; FEATURE:
; OTHER INFORMATION: Serine protease catalytic domain of hepsin (Heps)
; OTHER INFORMATION: homologous to similar domain in TADG-15
US-09-027-337-3

```

```

Query Match 19.5%; Score 528.5; DB 2; Length 256;
Best Local Similarity 41.2%; Pred. No. 9.2e-41;
Matches 101; Conservative 41; Mismatches 92; Indels 11; Gaps 3;
QY 255 RIVGCEGSLPGAMPQVSLHVNQVHVGSGSIITPEWITAAHCVEKPLNNPMHTAFAGI 314
Db 1 RIVGGRQTSILGRMPQVSLRYDGAHLGCGSLISGDMVLTAAHCPEERRRVLSRVRFAGA 60
QY 315 LROSEMFYAGAYQVEKYVSHPNY-----DSKTKNNDIALKLOKPLTFNDLVKPVCLPN 368
Db 61 VAOASP-HGDLGVQAVVYHGGYLPFRDPNSEENSDIALVHLSSPLPLEYIOPVCLPA 119
QY 369 PGMMLQPEQLCWMISGMATEEKGKTSVYLNAAKYLLETQRCNSRHYVDNLITPAMICAG 428
Db 120 AGQALVYDGKICTVYGMWNTQFYGGQAGVLOEARPILISNDVCSNADPFGNIDKFKMFCAG 179

```


QY 368 NPGMWLOEOLCWSMGATEEKGKTSSEVLNAKAVLLIETORCSRRYYDNLITPAMICA 427
Db 116 SKGRBNVITTCOWWYMGVYRKLDRKIQNTLQKAKIPLVTNECCQARY-RGKHITFKMICA 174
QY 428 GFLGNDVSCGDSGGLVTSKNNIMWLIGDTSMGSCAKAYRPGVYGNVMTMIRQ 487
Db 175 GYREGKACKAGDSGGLSCAKHNEVMHLVGTSMGSCAQRERPGVYTNVEYVDMLEK 234
QY 488 MRA 490
Db 235 TQA 237
RESULT 14
US-08-944-483-67
Sequence 67, Application US/08944483
Patent No. 6232456
GENERAL INFORMATION:
APPLICANT: COHEN, MAURICE
APPLICANT: COLPITTS, TRACEY L.
APPLICANT: FRIEDMAN, PAULA N.
APPLICANT: GRANADOS, EDWARD N.
APPLICANT: KLASS, MICHAEL R.
APPLICANT: RUSSELL, JOHN C.
APPLICANT: STEWART, KENT D.
APPLICANT: STROUPE, STEVEN D.
TITLE OF INVENTION: NOVEL SERINE PROTEASE REAGENTS
TITLE OF INVENTION: AND METHODS USEFUL FOR DETECTING AND TREATING DISEASES
NUMBER OF SEQUENCES: 76
CORRESPONDENCE ADDRESS:
ADDRESSEE: Abbott Laboratories
STREET: 100 Abbott Park Road
CITY: Abbott Park
STATE: IL
COUNTRY: USA
ZIP: 60064-3500
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/944,483
FILING DATE:
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Becker, Cheryl L.
REGISTRATION NUMBER: 35,441
REFERENCE/DOCKET NUMBER: 6183 US 01
TELECOMMUNICATION INFORMATION:
TELEPHONE: 847/935-1729
TELEFAX: 847/938-2623
TELEX:
INFORMATION FOR SEQ ID NO: 67:
SEQUENCE CHARACTERISTICS:
LENGTH: 255 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: No. 6232456e
US-08-944-483-67

Query Match 19.3%; Score 523.5; DB 4; Length 255;
Best Local Similarity 41.0%; Pred. No. 2.6e-40;
Matches 100; Conservative 41; Mismatches 92; Indels 11; Gaps 3;
QY 256 IVGGSALPGAMPQVSLHVONVHVCSSITTPETVTAACVKEPLNPNHMTAFACIL 315
||||| : ||||||| : ||||||| : ||||||| : ||| :

Db 1 IVGGRDTSIGRMPQVSLRYDGAHLGGSLLSGDVLTAAHCFFERNRVLNRRVFAGAV 60
QY 316 ROSFMETGAGVYEVKIVSHPNY-----DSKTKNDIALMKLQKPLTFENDLVKPCCLPNP 369
Db 61 AQAAP-HGLQLGVQAVYVHGGLPFRDRNSSENSNDILVHLSPLPLETIQVCLPAA 119
QY 370 GMLLOEOLCWSMGATEEKGKTSSEVLNAKAVLLIETORCSRRYYDNLITPAMICAGF 429
Db 120 GOALVDGKICFTVYGMNTQYVQAGVLOEARVPIISNDVCGADFGYGNQIKPKMECAGY 179
QY 430 LOGNDVSCGDSGGLV-----TSKNNIMWLIGDTSMGSCAKAYRPGVYGNVMTMIRY 485
Db 180 PEGGIDACQDSDGPGFVCDISIRTPRRRLCGIVSWGTCALAKRPGVYTVSDREMIIF 239
QY 486 ROMR 489
Db 240 QAIR 243

RESULT 15
US-08-807-151-3
Sequence 3, Application US/08807151
Patent No. 6043033
GENERAL INFORMATION:
APPLICANT: Bandman, Olga
APPLICANT: Lal, Preeti
TITLE OF INVENTION: NOVEL HUMAN PROSTATE-ASSOCIATED
TITLE OF INVENTION: PROTEASE
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: US
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/807,151
FILING DATE: Filed Herewith
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PE-0227 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-855-0555
TELEFAX: 415-845-4166
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 235 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: Genbank
CLONE: 416132
US-08-807-151-3

Query Match 19.2%; Score 521.5; DB 3; Length 235;
Best Local Similarity 42.3%; Pred. No. 3.6e-40;
Matches 99; Conservative 44; Mismatches 82; Indels 9; Gaps 3;
QY 256 IVGGSALPGAMPQVSLHVONVHVCSSITTPETVTAACVKEPLNPNHMTAFACIL 315
||||| : ||||||| : ||| : ||| : ||| : ||| : ||| :

Db 1 IVGSDSREGAMPWVVALYEDDOVCASLVSRDMLVSAHCVYGRNMEPSKKAVALGLH 60
QY 316 ROSFMEYGAGYVE-----KVISHPRYDSKTKNNNDIALMKIQKPLFENDLVKPVCLPNPG 370
Db 61 MASNL---TSPQIETRLIDQIVINPHYNNRRKNNDIAMHLEKKNVYTDYIOPICLPEEN 117
QY 371 MMLOPEOLCWMISGWGATEEKGKTSEVLNNAKVLLETOQNSRYYVDNLITPAMICAGFL 430
Db 118 QYFPPGRICSTAGWGALLIYGSTADYLOEADYPLLSNEKCOQOMPEYN-ITENMYCAGYE 176
QY 431 QGNVDSGCGDSGGLVTSKNNIMWMLIGDTSWGSCKAKAYRPGVYGNVWVFTDNI 484
Db 177 AGGVDSGCGDSGGLVTSKNNIMWMLIGDTSWGSCKAKAYRPGVYGNVWVFTDNI 230

Search completed: September 26, 2001, 09:22:08
Job time: 63 sec

THIS PAGE BLANK (USPTO)